

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 16:25:52 ; Search time 623 Seconds
(without alignments)
10334.612 Million cell updates/sec

Title: US-09-923-684-2
Perfect score: 2859
Sequence: 1 actcactataggctcgagc.....gtctcaaaaaaaaaaaaaa 2859

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2859	100.0	2859	24	Human single minde
2	1661.6	58.1	3920	24	Human single minde
3	1625.6	56.9	3885	24	Prostate cancer-as
4	1369	47.9	1464	24	Prostate cancer-as
5	1323.4	46.3	3614	18	Murine SIM (single
6	704.6	24.6	4557	23	DNA encoding novel
7	465.6	16.3	472	24	Unigene #2 used in
8	452	15.8	1905	23	Drosophila melanog
9	231.4	8.1	238	18	Partial human SIM

10	230.6	8.1	711	22	AAK82492	Human immune/haema
11	230.6	8.1	711	22	AAK82495	Human immune/haema
12	230.6	8.1	50000	24	ABK98944	Human asthma-assoc
13	228.8	8.0	11746	22	AAK73167	Human immune/haema
14	228.8	8.0	11746	24	ABK69935	Human secreted pro
15	225.2	7.9	10612	22	AAK36946	Human musculoskele
16	225.2	7.9	10614	22	AAK36948	Human musculoskele
17	224.4	7.8	11337	22	AAK36286	Human polynucleoti
18	224.4	7.8	7818	22	AAK51950	Human immune/haema
19	224	7.8	42519	22	AAK81318	Human reproductive
20	223.8	7.8	1743	22	AAK03368	Human breast or ov
21	223.8	7.8	1743	22	AAI62586	Human reproductive
22	223.8	7.8	1746	22	AAK03369	Human breast or ov
23	223.8	7.8	1746	22	AAI62587	Human pancreatic c
24	223.8	7.8	8676	22	AAK07159	Human cancer relat
25	223.8	7.8	31931	22	ABK07160	Human musculoskele
26	223	7.8	606	22	ABN62170	Genomic sequence #
27	222.8	7.8	8207	22	AAK36305	Human ovarian and
28	222.8	7.8	8207	22	AAK39844	Human reproductive
29	222.8	7.8	8207	22	AAK90215	Human ovarian and
30	222.6	7.8	32134	22	ABK07813	Human reproductive
31	222.6	7.8	32134	22	AAK03615	Human immune/haema
32	222.6	7.8	32191	22	ABK07814	Human immune/haema
33	222.6	7.8	32191	22	AAK03616	Human immune/haema
34	222.4	7.8	4125	22	AAK84260	Human immune/haema
35	222.4	7.8	4129	22	AAK84261	Human immune/haema
36	222.4	7.8	43599	24	ABK84242	Human cDNA differe
37	222.2	7.8	6784	22	AAK69691	Human immune/haema
38	222.2	7.8	6784	22	AAK69693	Human immune/haema
39	222.2	7.8	7240	22	AAH50570	Insulin receptor g
40	222.2	7.8	7240	22	AAH31170	Human insulin rece
41	222.2	7.8	7240	22	AAH31267	Human insulin rece
42	222.2	7.8	7240	22	AAH04467	Human insulin rece
43	222.2	7.8	7240	22	ABK07275	Human insulin rece
44	221.8	7.8	1301	21	AAH51627	Human g35018 cDNA
45	221.8	7.8	1386	21	AAH51631	Human g35018 cDNA

ALIGNMENTS

RESULT 1
AAD30523
ID AAD30523 standard; DNA; 2859 BP.

XX AAD30523;

AC AAD30523;

DT 31-MAY-2002 (first entry)

XX Human single minded homologue 2 (SIM2) gene, short form.

XX Human: single minded homologue 2; SIM2; colon cancer; prostate cancer;

XX pancreas cancer; chromosome 21q22.2; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 93..1805

XX /*tag= a

XX /product= "Human SIM2 protein, short form"

XX WO200212565-A1.

XX 14-FEB-2002.

XX 06-AUG-2001; 2001WO-US24781.

XX 04-AUG-2000; 2000US-223531P.

XX 22-DEC-2000; 2000US-257965P.

XX (UYFL) UNIV FLORIDA ATLANTIC.

XX Narayanan R;

XX WPI; 2002-227169/28.
DR P-PSDB; AAE19388.
XX
PT Detecting cancer in tissue sample, involves providing tissue sample,
PT and analyzing tissue sample for presence of single minded homolog 2
PT marker
XX
XX Claim 9; Page 55-56; 60pp; English.
XX
XX The invention relates to a method of detecting cancer in a tissue
CC sample. The method involves providing the tissue sample and analyzing
CC the tissue sample for the presence of a single minded homolog 2 (SIM2)
CC nucleic acid or protein, where the presence of the SIM2 marker in the
CC tissue sample indicates that the tissue sample contains cancer. The
CC tissue sample is useful for identifying compounds that modulate expression of
CC SIM2 gene in a cell. It is useful for detecting and reducing the growth
CC of cancer such as colon, prostate and pancreas cancer in an animal or
CC mammal. The present sequence is human SIM2 gene, short form. SIM2 gene
CC is located on chromosome 21q22.2.
XX
XX Sequence 2859 BP; 701 A; 803 C; 778 G; 577 T; 0 other;
SQ
Query Match 100.0%; Score 2859; DB 24; Length 2859;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACTCACTATAGGGCTCGAGCGCGCGCGGGGAGGTGGGGCTCCCGGGCTGGAGCAGC 60
DB 1 ACTCACTATAGGGCTCGAGCGCGCGCGGGGAGGTGGGGCTCCCGGGCTGGAGCAGC 60
QY 61 GCGGGTCTAATATGCGCGGAGCGCGGCGGCGGATGAAGAGAGTCCAAAGATGCGGCCA 120
DB 61 GCGGGTCTAATATGCGCGGAGCGCGGCGGCGGATGAAGAGAGTCCAAAGATGCGGCCA 120
QY 121 AGACAGGAGGAGAGAGAAATGGCGAGTTTACGAGCTTGCCAAAGTGTCCCGCTGC 180
DB 121 AGACAGGAGGAGAGAAATGGCGAGTTTACGAGCTTGCCAAAGTGTCCCGCTGC 180
QY 181 CGTGGCGCATCTTCCGAGCTGGACAAAGCTCCATATCGCTCCACAGAGTACC 240
DB 181 CGTGGCGCATCTTCCGAGCTGGACAAAGCTCCATATCGCTCCACAGAGTACC 240
QY 241 TGAAGATGCGCGCTTCCCGAGAGTTTAGAGACGCGTGGGGAGCAGCGCGCG 300
DB 241 TGAAGATGCGCGCTTCCCGAGAGTTTAGAGACGCGTGGGGAGCAGCGCGCG 300
QY 301 CCGGGCCCTGGAGCGCGTCCCAAGAGCTGGGATCGCAGCTTGTGATG 360
DB 301 CCGGGCCCTGGAGCGCGTCCCAAGAGCTGGGATCGCAGCTTGTGATG 360
QY 361 GATTTGTTTGTGTAGCATCTGATGGCAAAATCATGTATATCCGAGACCGCTTCTG 420
DB 361 GATTTGTTTGTGTAGCATCTGATGGCAAAATCATGTATATCCGAGACCGCTTCTG 420
QY 421 TCCATTTAGGCTTATCCAGGTGAGCTCAGCGGCACAGTATTATGAATACATCATC 480
DB 421 TCCATTTAGGCTTATCCAGGTGAGCTCAGCGGCACAGTATTATGAATACATCATC 480
QY 481 CTTCTGACACGATGAGATGACCGCTGTCTCAGCGGCCACAGCGCGCTGCACACACC 540
DB 481 CTTCTGACACGATGAGATGACCGCTGTCTCAGCGGCCACAGCGCGCTGCACACACC 540
QY 541 TGCTCCAAAGATGATGAGATGAGAGGTGTTCTTTCTCGAATGATGCTTGGCGA 600
DB 541 TGCTCCAAAGATGATGAGATGAGAGGTGTTCTTTCTCGAATGATGCTTGGCGA 600
QY 601 AAAGAACCGCGGCTGACCTGCAGCGGATACAGGTTCATCCACTGCGAGTGGCTTCTGA 660
DB 601 AAAGAACCGCGGCTGACCTGCAGCGGATACAGGTTCATCCACTGCGAGTGGCTTCTGA 660
QY 661 AGATCAGGAGTATATGCTGGACATGTCCCTGTAGCAGTCTCTGTACAGATTGTGGGC 720
DB 661 AGATCAGGAGTATATGCTGGACATGTCCCTGTAGCAGTCTCTGTACAGATTGTGGGC 720

DB 661 AGATCAGGAGTATATGCTGGACATGTCCCTGTAGCAGTCTCTGTACAGATTGTGGGC 720
QY 721 TGGTGGCGGTGGCGCAGTCCGCTGCCACCCAGTGCATCACCAGATCAAGCTGTACAGTA 780
DB 721 TGGTGGCGGTGGCGCAGTCCGCTGCCACCCAGTGCATCACCAGATCAAGCTGTACAGTA 780
QY 781 ACATCTTTCATGTTTCCAGGGCCAGCCTTGACCTGAAGCTGATATCTCTGATTCAGGGTGA 840
DB 781 ACATCTTTCATGTTTCCAGGGCCAGCCTTGACCTGAAGCTGATATCTCTGATTCAGGGTGA 840
QY 841 CCGAGGTGACGGGTTTACGAGCGCGGAGGACCTGTGAGAGAACCTTATACCATCACGTGC 900
DB 841 CCGAGGTGACGGGTTTACGAGCGCGGAGGACCTGTGAGAGAACCTTATACCATCACGTGC 900
QY 901 ACGGCTGCGACCTGTTTCCACCTCCCTACGACACACCTCTCTGTTGTTGAGGGCCAGG 960
DB 901 ACGGCTGCGACCTGTTTCCACCTCCCTACGACACACCTCTCTGTTGTTGAGGGCCAGG 960
QY 961 TCACCAACCAAGTACTACCGGCTGCTTCCAAAGCGGGGGCTGGGTGGTGGTGGCAGAGCT 1020
DB 961 TCACCAACCAAGTACTACCGGCTGCTTCCAAAGCGGGGGCTGGGTGGTGGTGGCAGAGCT 1020
QY 1021 ACGCCACCGTGTGTCACAAAGCGCGCTCGTCCCGGCCCTGCTGATCGTGTAGTCAATT 1080
DB 1021 ACGCCACCGTGTGTCACAAAGCGCGCTCGTCCCGGCCCTGCTGATCGTGTAGTCAATT 1080
QY 1081 ATGACTCAGGAGATGATATACAGGAATTCAGCTGCTCCCTGGAGCAGTGTCCACTG 1140
DB 1081 ATGACTCAGGAGATGATATACAGGAATTCAGCTGCTCCCTGGAGCAGTGTCCACTG 1140
QY 1141 CCAAGTCCCAGGACTCTCTGGAGGACCGCTTGTCTACTCACAAGAAATAGAAATTAG 1200
DB 1141 CCAAGTCCCAGGACTCTCTGGAGGACCGCTTGTCTACTCACAAGAAATAGAAATTAG 1200
QY 1201 TGAACCCCAAAATACCAAGATGAAGCAAACTGGAATGCGGCCAGCTCGGAAATGAGAGC 1260
DB 1201 TGAACCCCAAAATACCAAGATGAAGCAAACTGGAATGCGGCCAGCTCGGAAATGAGAGC 1260
QY 1261 AATACAGCTCATTTCCAAATGGAATGGAATGCGGCCAGCTCGGAAATGAGAGC 1320
DB 1261 AATACAGCTCATTTCCAAATGGAATGGAATGCGGCCAGCTCGGAAATGAGAGC 1320
QY 1321 GTCCCGCTTGCAGGCGTGTCTCTCCAGAACTGCAAGCACTCAGAAAGCAGTGAAC 1380
DB 1321 GTCCCGCTTGCAGGCGTGTCTCTCCAGAACTGCAAGCACTCAGAAAGCAGTGAAC 1380
QY 1381 TTCTGTACAGCCATCTTACAGCTGCGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1440
DB 1381 TTCTGTACAGCCATCTTACAGCTGCGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1440
QY 1441 ACTCTCAGCTTCTTACAGCAGCAAAAGCAATGTTGCGGCCAAGTTCGGCAGCGCCCAAG 1500
DB 1441 ACTCTCAGCTTCTTACAGCAGCAAAAGCAATGTTGCGGCCAAGTTCGGCAGCGCCCAAG 1500
QY 1501 GATCCCGTGTGAGTGGCAGCTTTTCTTCTGAGCACAATGCCAGCCAGCGGTGAATGCC 1560
DB 1501 GATCCCGTGTGAGTGGCAGCTTTTCTTCTGAGCACAATGCCAGCCAGCGGTGAATGCC 1560
QY 1561 AGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTCTCTCCAGCTAAAATCCTCCAG 1620
DB 1561 AGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTCTCTCCAGCTAAAATCCTCCAG 1620
QY 1621 AGCCACCGCGACACTGCTAGGCACAGCTGTCGACAGCTTACGAAGTGGGTGAGTGC 1680
DB 1621 AGCCACCGCGACACTGCTAGGCACAGCTGTCGACAGCTTACGAAGTGGGTGAGTGC 1680
QY 1681 TGCTCTGTGGGAGGTGGGAGGACTGCGCACCGCGGGGAGCGGAGCAGCCATGCGGTG 1740
DB 1681 TGCTCTGTGGGAGGTGGGAGGACTGCGCACCGCGGGGAGCGGAGCAGCCATGCGGTG 1740
QY 1741 GGTGGCAGATGGAGACAGAACCTCTTGGGCAAACTTGGCCCTCTTTCTGCTCTTA 1800
DB 1741 GGTGGCAGATGGAGACAGAACCTCTTGGGCAAACTTGGCCCTCTTTCTGCTCTTA 1800

CC The present invention relates to methods of detecting a prostate
CC cancer-associated transcript in a cell from a patient. The method
CC comprises contacting a biological sample from the patient with
CC prostate cancer-associated polynucleotides (designated PC genes) that
CC selectively hybridize to a sequence that is at least 80% identical
CC to them. The prostate cancer-associated polynucleotide sequences
CC are differentially expressed in prostate tumour tissue or in
CC prostate cancer and are derived from the tissues of various
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC The methods of the invention are useful for diagnosing and treating
CC prostate cancer in mammals. The prostate cancer-associated genes are
CC useful for diagnosing or treating prostate cancer, as well as for
CC identifying modulators of prostate cancer or agents that inhibit
CC prostate cancer. The nucleic acid sequences are particularly useful
CC in gene therapy, as a vaccine or in antisense applications.
CC ABK92115-ABK92363 represent prostate cancer-associated polynucleotide
CC sequences.
XX

Sequence 3885 BP; 916 A; 1125 C; 985 G; 859 T; 0 other;
Query Match 56.9%; Score 1625.6; DB 24; Length 3885;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1628; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 37 GGGGCTCCGGCCGTGGAGACGCGCGGTCTAATATGCCGGAGCGCGGATGA 96
DB 1 GGGGCTCCGGCGGCTGGAGACGCGCGGTCTAATATGCCGGAGCGCGGATGA 60
QY 97 AGGAGAAGTCCAAAGATCGGCAAGACGAGGAGGAGAAAGTGGCGAGTTACG 156
DB 61 AGGAGAAGTCCAAAGATCGGCGCAAGACGAGGAGGAGAAAGTGGCGAGTTACG 120
QY 157 ACCTTCCCAAGTGTCCGGTGGCGGCTGCGGCATCATCTTCAGCTGGACAAAGCGTCCA 216
DB 121 ACCTTGGCAAGTGTCCCGTGGCGTGGCCATCATCTTCAGCTGGACAAAGCGTCCA 180
QY 217 TCATCCGCTCACCACGAGTACCTGAAGATGCGCGCGTCTCCCGAAGGTTAGGAG 276
DB 181 TCATCCGCTCACCACGAGTACCTGAAGATGCGCGCGTCTCCCGAAGGTTAGGAG 240
QY 277 AGCGTGGGACAGCGGAT 336
DB 241 AGCGTGGGACAGCGGAT 300
QY 337 CGCACTTGTCTCAGACGTTGGATGATTTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 396
DB 301 CGCACTTGTCTCAGACGTTGGATGATTTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 360
QY 397 TGTATATATCGAGACGCGCTTCTGTCCATTTAGGCTTATCCAGGTGGAGCTCAGCGGA 456
DB 361 TGTATATATCGAGACGCGCTTCTGTCCATTTAGGCTTATCCAGGTGGAGCTCAGCGGA 420
QY 457 ACAGTATTTATGAATACATCCCTTCTGACACGATGAGATGACCGGTGCTCAGCG 516
DB 421 ACAGTATTTATGAATACATCCCTTCTGACACGATGAGATGAGAGGTGCTTCTTC 480
QY 517 CCACACGCGCTGCACACACCTGCTCCAGAGTATGAGATGAGAGGTGCTTCTTC 576
DB 481 CCACACGCGCTGCACACACCTGCTCCAGAGTATGAGATGAGAGGTGCTTCTTC 540
QY 577 TTCGAATGAATGTCTTGGCGAAAGAAAGCGCGGCGCTGACCTGCAGCGGATACAAGG 636
DB 541 TTCGAATGAATGTCTTGGCGAAAGAAAGCGCGGCGCTGACCTGCAGCGGATACAAGG 600
QY 637 TCATCCATGTCAGTGGCTACTTGAAGATCAGCAGTATATGTCGACATGTCCTGTACG 696
DB 601 TCATCCATGTCAGTGGCTACTTGAAGATCAGCAGTATATGTCGACATGTCCTGTACG 660
QY 697 ACTCTCTACAGATTTGGGGCTGGTGGCGTGGCGAGTGCCTGCGCACCGAGTGCCA 756
DB 661 ACTCTCTACAGATTTGGGGCTGGTGGCGTGGCGAGTGCCTGCGCACCGAGTGCCA 720
QY 757 TCACCGAGATCAAGCTGTACATACATGTTTCATGTTGAGGGCCAGCGCTTGACCTGAAGC 816

DB 721 TCACCGAGATCAAGCTGTACAGTAACTGTTTCATGTTTCAGGGCCAGCGCTTGACCTGAGC 780
QY 817 TGATATTTCTGATTTCCAGGGTGACCGAGGTGAGCGGTTAGGAGCCGAGGACCTGATCG 876
DB 781 TGATATTTCTGATTTCCAGGGTGACCGAGGTGACCGGTTACGAGCCGAGGACCTGATCG 840
QY 877 AGAAGACCTATACCATCATCGTGCACGGCTGCGACGCTGTTCCACCTCCCGTACGACACC 936
DB 841 AGAAGACCTATACCATCATCGTGCACGGCTGCGACGCTGTTCCACCTCCCGTACGACACC 900
QY 937 ACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTACCGGCTGCTGTCCAGCGGG 996
DB 901 ACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTACCGGCTGCTGTCCAGCGGG 960
QY 997 GCGGCTGGGTGGTGGTGCAGAGTACGCCACCTGGTGCACACACCGCTCGTCCCGGC 1056
DB 961 GCGGCTGGGTGGTGGTGCAGAGTACGCCACCTGGTGCACACACCGCTCGTCCCGGC 1020
QY 1057 CCACCTGCATCGTGAAGTGTCAATTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 1116
DB 1021 CCACCTGCATCGTGAAGTGTCAATTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 1080
QY 1117 TGTCTCTGGAGCAGGTGTCCACTGCGCAAGTCCGAGGACTCTCTGGAGGACCGCTTGTCTTA 1176
DB 1081 TGTCTCTGGAGCAGGTGTCCACTGCGCAAGTCCGAGGACTCTCTGGAGGACCGCTTGTCTTA 1140
QY 1177 CCTCACAAGAACTAGGAAATTTAGTGAACCCCAAAATACCAAGATGAAGAAAGCTGA 1236
DB 1141 CCTCACAAGAACTAGGAAATTTAGTGAACCCCAAAATACCAAGATGAAGAAAGCTGA 1200
QY 1237 GAACAAACCTTACCCCCCAGCAATACAGCTCAATTCCTGAGGACCAAACTGGAATGCG 1296
DB 1201 GAACAAACCTTACCCCCCAGCAATACAGCTCAATTCCTGAGGACCAAACTGGAATGCG 1260
QY 1297 GCAGCTCGGAACCTGGAGAGCGAGTCCCGCTGCAAGCGCTGCTGCTCCTCAGAACTGC 1356
DB 1261 GCAGCTCGGAACCTGGAGAGCGAGTCCCGCTGCAAGCGCTGCTGCTCCTCAGAACTGC 1320
QY 1357 AGCCCACTCAGAAAGCAGTGAACCTTCTGTACAGCCATCTTACAGCTGCGCTTCTCCT 1416
DB 1321 AGCCCACTCAGAAAGCAGTGAACCTTCTGTACAGCCATCTTACAGCTGCGCTTCTCCT 1380
QY 1417 ACCATTATGACACTTCCCTCTGACACTCTCACTTCTTACAGCAGCAAAAGCCAAATGTTGC 1476
DB 1381 ACCATTATGACACTTCCCTCTGACACTCTCACTTCTTACAGCAGCAAAAGCCAAATGTTGC 1440
QY 1477 CGGCAAGTTCGGCAGCGCCCAAGGATCCCTTGTGAGGTGGCAGCGCTTTTCTCTGAGCA 1536
DB 1441 CGGCAAGTTCGGCAGCGCCCAAGGATCCCTTGTGAGGTGGCAGCGCTTTTCTCTGAGCA 1500
QY 1537 CAATGCCAGCCAGCGGTGAATGCGAGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCT 1596
DB 1501 CACTGCCAGCCAGCGGTGAATGCGAGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCT 1560
QY 1597 CGTCTCCAGCTAAAATCTCCAGAGCCACCGGCAACTGCTAGGACAGCGCTGCTGTC 1656
DB 1561 CGTCTCCAGCTAAAATCTCCAGAGCCACCGGCAACTGCTAGGACAGCGCTGCTGTC 1620
QY 1657 CAAGCTACGAG 1668
DB 1621 CAAGCTACGAG 1632

RESULT 4
ABK92250
ID ABK92250 standard; DNA; 1464 BP.
XX
AC ABK92250;
XX
DT 15-AUG-2002 (first entry)
XX
DE Prostate cancer-associated DNA sequence #136.

Db 1260 GCCAGCCAGGGTGAATGCCAGTGGCATATATGCCAACCCCTAGTGCCTAGCAGCTCGTC 1319
Qy 1601 TCCAGCTAAATTCCTCCAGAGCCACCGCGGAACACTGCTAGGCACAGCCTGGTGCCCAAG 1660
Db 1320 TCCAGCTAAATTCCTCCAGAGCCACCGCGGAACACTGCTAGGCACAGCCTGGTGCCCAAG 1379
Qy 1661 CTACGAAGG 1669
Db 1380 CTACGAAGG 1388

RESULT 5

AAAT91883
ID AAAT91883 standard; cDNA to mRNA; 3614 BP.
AC AAAT91883;
XX
XX
XX 30-JAN-1998 (first entry)
DE Murine SIM (single-minded) gene product coding sequence.
XX
KW Mouse; SIM; single-minded; early development; central nervous system;
KW CNS; diagnosis; Down's syndrome; N-terminal; ds.
XX
OS Mus sp.

XX
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 504..2477
FT /*tag= a

JP09268199-A.

14-OCT-1997.

28-MAR-1996; 96JP-0074261.

29-JAN-1996; 96JP-0012952.

31-MAR-1995; 95JP-0076362.

(FUJI) FUJISAWA PHARM CO LTD.

WPI; 1997-554711/51.

P-PSDB; AAW30758.

Human or mouse single-minded gene - used in the diagnosis of Down's syndrome

Claim 5; Pages 12-13; 17pp; Japanese.

This is the murine SIM (single-minded) gene. The SIM gene appears to be of importance in controlling the early development of the central nervous system. In particular the gene, and its products are used in the diagnosis of Down's syndrome.

Sequence 3614 BP; 833 A; 1110 C; 952 G; 717 T; 2 other;

Query Match 46.3%; Score 1323.4; DB 18; Length 3614;
Best Local Similarity 88.2%; Pred. No. 0;
Matches 1451; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

Qy 27 CCGGGCAGGTGGGGCTCCCGGGCTGGAGCACGCGGGTCTAATATGCCGGAGCCGA 86
Db 438 CAGCCCAAGCGGGACTCCCGGGCTGGAGCCAGGAGTCTAATATGCCAGAGCCGA 497
Qy 87 GCGCGCATCAAGGAGAGTCCCAAGATGCGGCCAAGACAGAGGAGGAGGAAATGGC 146
Db 498 GCGCGCATCAAGGAGAGTCCCAAGATGCGGCCAAGACAGAGGAGGAGGAAATGGC 557
Qy 147 GAGTTTTAGAGCTTGCCCAAGCTGCTCCCGTGGCGTCCGATCATCTCCAGCTGGAC 206
Db 558 GAGTTTTAGAGCTTGCCCAAGCTGCTCCCGTGGCGTCCGATCATCTCCAGCTGGAC 617

Qy 207 AAGCGTCCCATCATCCGCTCACCAGAGCTACTCTGAAGATGCGGCCGCTTTCCTCCCGAA 266
Db 618 AAGCGTCCCATCATCCGCTCACCAGAGCTACTCTGAAGATGCGGCCGCTTTCCTCCCGAA 677
Qy 267 GGTTTAGAGACGCGTGGGACAGCGCGCGCGGCCCTGTGACCGGCTCGCCCAAG 326
Db 678 GGTCTAGGACGCGTGGGACAGCGCGCGCGGCCCTGTGACCGGCTGGCAAA 737
Qy 327 GAGCTGGGATCGCACTTGTGTCAGACTTTGGATGGATTTGTTTGTGTAGCATCTGAT 386
Db 738 GAGCTGGGATCGCACTTGTGTCAGACTTAGATGGATTTGTTTGTGTGCTGCTGAT 797
Qy 387 GGCAAAATCATGTATATATCCGAGACCGCTCTGTCCATTTAGGCTTATCCAGGTGAG 446
Db 798 GGCAAAATCATGTATATATGAGACCGCTCTGTCCATTTAGGCTTGTCCAGGTGAG 857
Qy 447 CTCACGGGCAACAGTATTTATGAATACATCCATCTTGTGACACGATGAGATGACCGT 506
Db 858 CTCACAGGCAACAGTATTTATGAATACATACATCTTGTGACACGATGAGATGACGCC 917
Qy 507 GTCTCACGGCCACCGCGCTGACACCGACCTGCTCCCAAGATATGATGAGAGG 566
Db 918 GTACTTACAGCCACCGCGCTCCACCATCACCTGCTCCCAAGATGAGATGAGAGGG 977
Qy 567 TCGTTCTTTCTTCCGAATGAATGTGTCTTGGCGAAAGAACGCGCGGCTGACCTGCGAC 626
Db 978 TCTTTCTTTCTTCCGAATGAATGTGTCTTGGCGAAAGAACGCGCGGCTGACATGCGAGT 1037
Qy 627 GGATACAGGTGATCCACTGAGTGGCTACTTGAAGATCAGGAGTATATCTGGACATG 686
Db 1038 GGATACAGGTGATCCACTGAGTGGCTACTTGAAGATCAGGAGTATATCTGGACATG 1097
Qy 687 TCCCTGTAGACTCTCTGCTTACCAGATTTGCGGGCTGTGGCGGCTGGCGGCTGCGCTGCCA 746
Db 1098 TCCCTGTAGACTCTCTGCTTACCAGATTTGCGGGCTGTGGCGGCTGGCGGCTGCGCTGCCA 1157
Qy 747 CCCAGTGCATCAGCGAGATCAAGCTGTACAGTAAACATGTTTCATGTTTTCAGGCGCAGCTT 806
Db 1158 CCCAGTGCATCAGCGAGATCAAGCTGTACAGTAAACATGTTTCATGTTTTCAGGCGCAGCTT 1217
Qy 807 GACCTGAAGCTGATATCTCTGATTCCAGGCTGACCGAGGTGACGGGTGACGAGCCCGCAG 866
Db 1218 GACCTGAAGCTGATATCTCTGATTCCAGGCTGACCGAGGTGACGAGCTTATGAGCCACA 1277
Qy 867 GACCTGATCGAAGAACCTATACCATCAGTGCACGCGCTCGCAGCTGTTTCCACCTCCGC 936
Db 1278 GACCTGATCGAAGAACCTATACCATCAGTGCACGCGCTCGCAGCTTATGAGCCACA 1337
Qy 927 TACGCACACACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTTACCAGGTGCTG 986
Db 1338 TATGCACACACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTTACCAGGTGCTG 1397
Qy 987 TCCAAGCGGGCGGCTGGGTGCGAGCTAGCCACCGCTGTCACACCGCTGTCACACACCGCC 1046
Db 1398 TCCAAGCGGGCGGCTGGGTGCGAGCTAGCCACCGCTGTCACACCGCTGTCACACACCGCC 1457
Qy 1047 TCGTCCCGCGCCACTCTCATCTGATGTCATATTTGTCACGAGATTTGAATACAAG 1106
Db 1458 TCTCCCGGCTCTCTGAGTGTCAATTTATGTTCTCACGGATGTTGAATACAAG 1517
Qy 1107 GAACTTCAGCTGCTCCCTGGAGCAGGTGTCCACTGCCAAGTCCAGGACTCTTGGAGGACC 1166
Db 1518 GAACTTCAGCTGCTCTGAGCAGGTATCCACCTCTAAGTCCAGGAGTCTTGGAGGACC 1577
Qy 1167 GCCTTGTCTACCTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG 1226
Db 1578 ACCTTGTCTACCTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG 1637
Qy 1227 ACAAACTGAGAAACCTTACCCCGCACAGCAATACAGCTCATTTCCAATGACAAA 1286
Db 1638 ACAAACTGAGAAACCTTACCCCGCACAGCAATACAGCTCATTTCCAATGACAAA 1697
Qy 1287 CTGGAATGCGGCGAGCTCGGAAACTGGAGAGCCAGTCCCGCTGCAAGGCGTCTCTCT 1346

example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 sequences represent sequences used in the exemplification of the present invention.

Query Match	8.1%	Score 230.6;	DB 22;	Length 711;
Best Local Similarity	80.2%;	Prod. No. 5.8e-50;		
Matches 283;	Conservative 0;	Mismatches 69;	Indels 1;	Gaps 1;
QY	2508	AGATGCGCTCAGTGCCTCAATTAATATGTGGGTGAAGTGCACATCAGGATTATGTGCCCCAGG	2567	
DB	59	AAATGCCAGAGGAACCTGCTAAAGGTGTGCACATGGCCATCAAGAAAGTATGTGACTGCG	118	
QY	2568	CCGGGCTCAGTGGCTCACACCTGTAATCCAGACACTTTGGGAGGCCAAGGTGGCGGATC	2627	
DB	119	CCGGGCGGAGTGGCTCAGCGCTGTAATCCAGCACTTTGGGAGGCCAGGTGGCGAGTTC	178	
QY	2628	ACCTGAGGTCAGGAGTTTGCACAAGCCT-GCCAAACAGCTGAACCCCATCTCCACTAA	2686	
DB	179	ACCTGAGGTTAGAAGTTTGAGCCAGCGCTGGGCCAACATGACAAACCCCGTCTCTACTAA	238	
QY	2687	AAATACAAAAATTAGTTGGGCATGGTGGTGAGCACCTGTATCCAGCTACTCTGGAGGC	2746	
DB	239	AGATACAAAAATATCTGGGTGCTGGTGGTGTGCACCTGTAATCCAGCTACTTGGGAGGC	298	
QY	2747	TGAGATAGGAGATCATTGAACCCGGGAGGTGGAGGTTCAGTGAGCTTAAGATCACATC	2806	
DB	299	TGAGGCAGAAGAAATCGCTTGAAACCCAGGAGGTGGAGGTTCAGTGAGCCCAAGATCACACC	358	
QY	2807	ACTGCACCTCCAGCCTGGGTAAACAGAGTGAGAGCTGTCTCAAAAAAAAAAAAAA	2859	
QY	2859	ACTGCACCTCCAGCCTGGGTAAACAGAGTGAGAGCTGTCTCAAAAAAAAAAAAAA	411	

RESULT 12
ABA98944
ID ABA98944 standard: DNA: 50000 BP:

XX ABA98944:

18-JUN-2002 (first entry)

XX DE Human asthma-associated gene AAGB genomic DNA #1.

XX Human; asthma; AACB; antiinflammatory; antiasthmatic; ARDS; COPD; COAD;
KW inflammatory disease; obstructive airways disease; dyspnea; emphysema;
KW adult respiratory distress syndrome; chronic bronchitis; eosinophil;
KW chronic obstructive pulmonary disease; pneumoconiosis;
KW chronic obstructive airways disease; gene; ds.

XX Homo sapiens.

XX
PN
WO200206312-A2.XX
PD
24-JAN-2002:

11-JUL-2001: 2001WO-EP08010.

13-III-2000: 2000US-0615247.

XX
PA / NOV 1 NOVARTIS AG.

PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

KKK

Whittaker PA;

WPT: 2002-195799/25.

Novel polypeptide encoded by disease associated gene, useful for treating an inflammatory or obstructive airways disease e.g. asthma

The sequence represents a human asthma associated gene (AAGB) genomic DNA sequence. The invention relates to a novel asthma-associated gene AAGB and the polypeptide encoded by AAGB. The polypeptide of the invention has anti-inflammatory and antiasthmatic activity and may have a use in gene therapy, or as a vaccine. The polypeptide, polynucleotide, antibody and antisense oligonucleotide of the invention (collectively referred to as agents) are useful for treating an inflammatory or obstructive airways disease. They are also useful for treating adult respiratory distress syndrome (ARDS), chronic obstructive pulmonary airways disease (COPD or COAD), including chronic bronchitis or dyspnea associated with it, emphysema, exacerbation of airways hyper-reactivity consequent to other drug therapy and pneumoconiosis. The agents are also useful in the treatment of eosinophil related disorders and asthma.

... F0000 BB: 14069 A: 9489 C: 10164 G: 16278 T: 0 other;

Query Match 8.1%; Score 230.6; DB 24; Length 50000;
Best Local Similarity 80.2%; Pred. No. 3.9e-49;
Matches 283; Conservative 0; Mismatches 69; Indels 1;

[illegible]

Db	40656	AAATCCCAAGGAAC	TGCTTAAAGGTGTGCACAGTGGCCATCAAGAAAGTAAGTGCACG	40711
QY	2568	CCGGGTCATG	GGTCAACACCTGTAATCCACGACTTTGGAGGSCAAGTGTGGCGGATC	2627

	CGGGCGCAGTGGCTACGCGTGTAATCCAGCACTTTGGAGGCCGAGGTGGGCAGGTC	40716	40775
Dd			

QY	2628	ACCTGAGGTCAGGAGTTTTCGCACAACGCT - GCCACAAGCTGAACCCCATCTCCACTAA	2688
bb	40776	ACCTGAGGTTAGAGTTTTCGACACGACCTGGGCACATGACAAACCCCGTCTCTACTAA	40835

DB	40776	ACCICGGGATACAGGCTTCTGGGATGTTGGGATGTTGGTGGACACTGTATCCAGCTACTCTCGAGGC	2746
QY	2687	AAATACAAAAATTAGTTGGGATGTTGGTGGACACTGTATCCAGCTACTCTCGAGGC	2746

Db

[illegible][illegible]

RESULT 13

AAK73167/C

AAK73167 standard; DNA; 11746 BP.

XX
XX

AAK73167;

06-NOV-2001 (first entry)

XX
00 NOV 1968
D1
CFO RD NO. 27070

DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27373
.....	

XX
XX
XX

Human; immune; haematopoietic; immune/haematopoietic

KW cytostatic; 9

XX

OS Homo sapiens.

XX.
33000157107-27

PN WO20015/182-A2.

XX
DD
09-AUG-2001

PD 09-AUG-2001.
YY

VV

PF 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-02114886.
PR 30-JUN-2000; 2000US-02115135.
PR 07-JUL-2000; 2000US-02116647.
PR 07-JUL-2000; 2000US-02116880.
PR 11-JUL-2000; 2000US-02117487.
PR 11-JUL-2000; 2000US-02117496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-02311242.
PR 08-SEP-2000; 2000US-02311243.
PR 08-SEP-2000; 2000US-02311244.
PR 08-SEP-2000; 2000US-02311413.
PR 08-SEP-2000; 2000US-02311414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-02321968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,

QY 2792 AGCTAAGATCACATCACTCACTCCAGCTCGGTAAACAGAGTGAGACTGTCTCAAAAAA 2851
DB 3606 AGCCAGATCGCACCTCACTCCAGCTCGGTAAACAGAGTGAGACTGTCTCAAAAAA 3547
QY 2852 AAAAAAA 2859
DB 3546 AAAAAAA 3539

RESULT 15
AAL36946/c
ID AAL36946 standard; DNA; 10612 BP;
XX AAL36946;
AC AAL36946;
XX 08-JAN-2002 (first entry)
DT XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3311.
XX
KW Cytostatic; Immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antifungal; antiparasitic;
KW vulnerary; anticonvulsant; antibacterial; antiinflammatory; antitumor;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.

OS Homo sapiens.
XX
XX WO200155367-A1.
PN
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01338.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225274.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 20:59:29 ; Search time 181 Seconds
(without alignments)
273.547 Million cell updates/sec

Title: US-09-923-684-12
Perfect score: 24
Sequence: 1 gagagcaagaagcacagcaagcc 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 693762

Minimum DB seq length: 0
Maximum DB seq length: 25

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	10	US-09-923-684-12
2	13.6	56.7	20	12	US-09-961-001-62
3	13.6	56.7	25	15	US-10-098-263B-19661
4	13.2	55.0	19	11	US-09-969-373-2893
5	13.2	55.0	25	12	US-09-992-665-108
6	13.2	55.0	25	15	US-10-098-263B-44300
7	13.2	55.0	25	15	US-10-098-263B-67339
8	13	54.2	22	15	US-10-121-393-13
9	13	54.2	25	15	US-10-098-263B-20997
10	13	54.2	25	15	US-10-098-263B-74151
11	12.8	53.3	20	12	US-09-104-750-32
12	12.8	53.3	25	12	US-09-754-853A-902
13	12.8	53.3	25	15	US-10-215-112-966
14	12.8	53.3	25	15	US-10-215-112-7212
15	12.8	53.3	25	15	US-10-098-263B-33827
16	12.6	52.5	20	12	US-09-771-933-112

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

c	17	12.6	52.5	21	12	US-09-771-933-154
	18	12.6	52.5	22	10	US-09-915-060-11
	19	12.6	52.5	22	10	US-09-915-060-45
	20	12.6	52.5	23	11	US-09-901-484A-143
	21	12.6	52.5	23	11	US-09-853-526-143
	22	12.6	52.5	25	11	US-09-827-998-1375
	23	12.6	52.5	25	11	US-09-827-998-1375
	24	12.6	52.5	25	11	US-09-827-998-1377
	25	12.6	52.5	25	11	US-09-827-998-1377
	26	12.6	52.5	25	11	US-09-827-998-1378
	27	12.6	52.5	25	11	US-09-827-998-1379
	28	12.6	52.5	25	11	US-09-827-998-1381
c	29	12.6	52.5	25	15	US-10-098-263B-80130
	30	12.4	51.7	25	10	US-09-866-108-14540
	31	12.4	51.7	25	10	US-09-866-108-14540
	32	12.4	51.7	25	10	US-09-866-108-14542
	33	12.4	51.7	25	10	US-09-866-108-14543
	34	12.4	51.7	25	10	US-09-853-688-31
c	35	12.4	51.7	25	15	US-10-098-263B-85390
	36	12.2	50.8	17	11	US-09-864-785-672
	37	12.2	50.8	19	12	US-09-864-636A-914
c	38	12.2	50.8	20	10	US-09-452-599-159
	39	12.2	50.8	20	12	US-09-915-814-154
c	40	12.2	50.8	21	15	US-10-239-804-45
	41	12.2	50.8	23	11	US-09-263-959-1262
c	42	12.2	50.8	23	12	US-09-864-636A-938
	43	12.2	50.8	24	11	US-09-817-014-9
	44	12.2	50.8	25	10	US-09-757-041-7
c	45	12.2	50.8	25	15	US-10-215-112-8929

ALIGNMENTS

RESULT 1
US-09-923-684-12
; Sequence 12, Application US/09923684
; Patent No. US20020081613A1
; GENERAL INFORMATION:
; APPLICANT: Narayanan, Ramaswamy
; TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER
; FILE REFERENCE: 6818-24
; CURRENT APPLICATION NUMBER: US/09/923.684
; CURRENT FILING DATE: 2001-09-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 24
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-923-684-12

Query Match 100.0%; Score 24; DB 10; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GAGAGCAAGAAGCACAGCAAGCC	24
Db	1	GAGAGCAAGAAGCACAGCAAGCC	24

RESULT 2
US-09-961-001-62
; Sequence 62, Application US/09961001
; Publication No. US20030109466A1
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF KSR EXPRESSION
; FILE REFERENCE: RTS-0280
; CURRENT APPLICATION NUMBER: US/09/961.001
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 87

```

; SEQ ID NO 62
; LENGTH: 20
; TYPE: DNA
; ORGANISM:
; FEATURE:
; OTHER INFO:
US-09-961-001-

```

Query Match 56.7%; Score 13.6; DB 12; Length 20;
Best Local Similarity 80.0%; Pred. No. 7.9e+03;
Matches 16; Conservative 0; Mismatches 4; Indels

QY
5 GCAAGAAAGCACAGCAAGCC 24
|| ||| ||||||| |

Dh
1 GCCAGATAGCACAGCCAGTC 20

RESULT 3

US-10-098-263B-19661/c
; Sequence 19661, Application US/10098263B
; Publication No. US20030104410A1

```

, GENERAL INFORMATION:
, APPLICANT: Mittman, Michael
, TITLE OF INVENTION: Human Microarray
, FILE REFERENCE: 3118.1
, CURRENT APPLICATION NUMBER: US/10/098,263B
, PRIOR FILING DATE: 2003-01-08
, CURRENT APPLICATION NUMBER: 60/276,759
, PRIOR FILING DATE: 2001-03-16
, NUMBER OF SEQ ID NOS: 131066
, SOFTWARE: Microarray Probe Sequence Listing
, SEQ ID NO 19661

```

SEQ	ID	NO
19661		

```

; LENGTH: 21
; TYPE: DNA

```

US-10-098-263B-19661

Query Match	Score 13.6;	DB 15;	Length 25;
Best Local Similarity	56.7%;	Pred. No. 8e+03;	Indels
	80.0%;	0. Mismatches	4.

QY
1 GAGAGCAAGAAAGCACAGCA 20
| | | | | | | | | |
ph
24 GCGAGCAAGTAAACACAGAA 5

RESULT 4

RESULTS 4
US-09-969-373-2893/c
; Sequence 2893, Application US/09969373
; Patent No US20020133852A1

```

, GENERAL INFORMATION:
, APPLICANT: Effertz, Roger J.
, APPLICANT: Haugbe, Brian M.
, TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
, FILE REFERENCE: 38-10(52679)A
, CURRENT APPLICATION NUMBER: US/09/969,373
, CURRENT FILING DATE: 2001-10-02
, PRIOR APPLICATION NUMBER: US 09/754,853
, PRIOR FILING DATE: 2001-01-05
, PRIOR APPLICATION NUMBER: US 09/760,427
, PRIOR FILING DATE: 2001-01-13
, PRIOR APPLICATION NUMBER: US 09/855,768
, PRIOR FILING DATE: 2001-05-15
, NUMBER OF SEQ ID NOS: 4593
, SEQ ID NO 2893
, LENGTH: 19

```

TYPE: DNA
ORGANISM: Glycine max

US-09-969-373-2893

Query Match 55.0%; Score 13.2; DB 11; Length 19;
Best Local Similarity 83.3%; Pred. No. 1.2e+04;
Matches 15: Conservative 0; Mismatches 3; Indels

QY 4 AGCAAGAAAGCACAGCAA 21
||||| ||

nb 19 AGCAAGCAAGCACATTA 2

RESULT 5

US-09-992-665-108
; Sequence 108, Application US/09992665
; Publication No. US2003092009A1

```

, GENERAL INFORMATION:
, APPLICANT: Kela Palm
, TITLE OF INVENTION: PROFILING TUMOR SPECIFIC MARKERS FOR THE
, TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF NEOPLASTIC DISEASE
, FILE REFERENCE: CEMINES-002A
, CURRENT APPLICATION NUMBER: US/09/992,665
, CURRENT FILING DATE: 2001-11-13
, PRIOR APPLICATION NUMBER: 60/249,508
, PRIOR FILING DATE: 2000-11-16
, NUMBER OF SEQ ID NOS: 380
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 108
, LENGTH: 25
, TYPE: DNA
, ORGANISM: Artificial Sequence
, FEATURE:
, OTHER INFORMATION: Probe
, US-09-992-665-108

```

Query Match 55.0%; Score 13.2; DB 12; Length 25;
Best Local Similarity 83.3%; Pred. No. 1.2e+04;
Matches 15; Conservative 0; Mismatches 3; Indels

Qy 3 GAGCAAGAAAGCACAGCA 20
||||| | ||| |||||

Db 2 GAGCATGGAAGAACAGCA 19

RESULT. T 6

US-10-098-263B-44300
; Sequence 44300, Application US/10098263B
Publication NO. US20030104410A1

```

> GENERAL INFORMATION:
> APPLICANT: Mittman, Michael
> TITLE OF INVENTION: Human Microarray
> FILE REFERENCE: 3118.1
> CURRENT APPLICATION NUMBER: US/10/098,263B
> CURRENT FILING DATE: 2003-01-08
> PRIOR APPLICATION NUMBER: 60/276,759
> PRIOR FILING DATE: 2001-03-16
> NUMBER OF SEQ ID NOS: 131066
> SOFTWARE: Microarray Probe Sequence Listing Generator v 1.1
> SEQ ID NO 44300
> LENGTH: 25
> TYPE: DNA
> ORGANISM: Homo sapien
> rs-10-098-263B-44300

```

Query Match	55.0%;	Score 13.2;	DB 15;	Length 25;
Best Local Similarity	83.3%;	Pred. No. 1.2e+04;		
Matches	15.	Conservative	0: Mismatches	3: Indels

QY
pb

5 GCAAGAAAGCACAGCAAG 22
| | | | | | | |
3 GGAAGAAAGGACATCAAG 20

RESULT 7

US-10-098-263B-67339/c
; Sequence 67339, Application US/10098263B
; Publication No. US20030104410A1
; GENERAL INFORMATION:
; APPLICANT: Mittman, Michael

; TITLE OF INVENTION: Human Microarray
; FILE REFERENCE: 3118.1
; CURRENT APPLICATION NUMBER: US/10/098,263B
; CURRENT FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 60/276,759
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 131066
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 67339
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-098-263B-67339

Query Match 55.0%; Score 13.2; DB 15; Length 25;
Best Local Similarity 83.3%; Pred. No. 1.2e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 CAGAAAGCAGCAGCAGC 23
DB 24 CGAGAAAGCATAGAAAGC 7

RESULT 8

US-10-121-393-13
; Sequence 13, Application US/10121393
; Publication No. US2003011563A1
; GENERAL INFORMATION:
; APPLICANT: Pals, Marie Salome Soares
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Balde, Aladjje
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES RELATING TO PAPAYA RIPENING
; CURRENT APPLICATION NUMBER: US/10/121,393
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/283,008
; PRIOR FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Amplification
US-10-121-393-13

Query Match 54.2%; Score 13; DB 15; Length 22;
Best Local Similarity 76.2%; Pred. No. 1.4e+04;
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCAGCAGCAG 22
DB 2 ACAGTAGAAGAACCATTCAG 22

RESULT 9

US-10-098-263B-20297/c
; Sequence 20297, Application US/10098263B
; Publication No. US20030104410A1
; GENERAL INFORMATION:
; APPLICANT: Mittman, Michael
; TITLE OF INVENTION: Human Microarray
; FILE REFERENCE: 3118.1
; CURRENT APPLICATION NUMBER: US/10/098,263B
; CURRENT FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 60/276,759
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 131066
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 20297
; LENGTH: 25

; TYPE: DNA
; ORGANISM: Homo sapien
US-10-098-263B-20297

Query Match 54.2%; Score 13; DB 15; Length 25;
Best Local Similarity 76.2%; Pred. No. 1.5e+04;
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCAGCAGCAG 22
DB 22 AGAGCAAGTAGTCACGAAAG 2

RESULT 10

US-10-098-263B-74151
; Sequence 74151, Application US/10098263B
; Publication No. US20030104410A1
; GENERAL INFORMATION:
; APPLICANT: Mittman, Michael
; TITLE OF INVENTION: Human Microarray
; FILE REFERENCE: 3118.1
; CURRENT APPLICATION NUMBER: US/10/098,263B
; CURRENT FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 60/276,759
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 131066
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 74151
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-098-263B-74151

Query Match 54.2%; Score 13; DB 15; Length 25;
Best Local Similarity 76.2%; Pred. No. 1.5e+04;
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCAGCAGCAG 22
DB 3 AGAGCAAGACACACAGAGAG 23

RESULT 11

US-09-104-750-32
; Sequence 32, Application US/09104750
; Publication No. US20030104364A1
; GENERAL INFORMATION:
; APPLICANT: Russell, John
; APPLICANT: Colpitts, Tracey
; TITLE OF INVENTION: Methods and Reagents Useful
; TITLE OF INVENTION: for Detecting Disease of the Breast
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,750
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 6121.US.01
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L

```

;
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6121.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-104-750-32
Query Match 53.3%; Score 12.8; DB 12; Length 20;
Best Local Similarity 87.5%; Pred. No. 1.7e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 CAAGCAAGCAGCAGCAA 21
   |||| ||||| ||||
Db 3 CAAGCAAGCAGCAGCAA 18

RESULT 12
US-09-754-853A-902
; Sequence 902, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Farnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 902
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 318013_region_A3_453879_15_Reverse_Primer_Seq
US-09-754-853A-902

Query Match 53.3%; Score 12.8; DB 12; Length 25;
Best Local Similarity 87.5%; Pred. No. 1.8e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCACA 17
   ||| ||||| |||
Db 10 AGAACAAGAAAGGACA 25

RESULT 13
US-10-215-112-966/c
; Sequence 966, Application US/10215112
; Publication No. US20030082596A1
; GENERAL INFORMATION:
; APPLICANT: Michael Mittmann
; TITLE OF INVENTION: Method of Genetic Analysis of Probes:
; FILE REFERENCE: 3119
; CURRENT APPLICATION NUMBER: US/10/215,112
; CURRENT FILING DATE: 2002-08-08
; NUMBER OF SEQ ID NOS: 14936
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 966
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE: DNA
; OTHER INFORMATION: Sequence Listing Generator V 1.1
US-10-098-263B-33827

Query Match 53.3%; Score 12.8; DB 15; Length 25;
Best Local Similarity 87.5%; Pred. No. 1.8e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 AAGAAAGCAGCAGCAAG 22
   ||||| ||||| |
Db 2 AAGAAAGTACAGCATG 17

US-10-215-112-7212
; Sequence 7212, Application US/10215112
; Publication No. US20030082596A1
; GENERAL INFORMATION:
; APPLICANT: Michael Mittmann
; TITLE OF INVENTION: Method of Genetic Analysis of Probes:
; FILE REFERENCE: 3119
; CURRENT APPLICATION NUMBER: US/10/215,112
; CURRENT FILING DATE: 2002-08-08
; NUMBER OF SEQ ID NOS: 14936
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7212
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-10-215-112-7212

Query Match 53.3%; Score 12.8; DB 15; Length 25;
Best Local Similarity 70.8%; Pred. No. 1.8e+04;
Matches 17; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAGAGCAAGAAAGCAGCAGGCC 24
   ||| |||| | |||| ||||
Db 25 GAGCGGAAGCATGAGCAGCTAGCC 2

US-10-098-263B-33827
; Sequence 33827, Application US/10098263B
; Publication No. US20030104410A1
; GENERAL INFORMATION:
; APPLICANT: Mittman, Michael
; TITLE OF INVENTION: Human Microarray
; FILE REFERENCE: 3118-1
; CURRENT APPLICATION NUMBER: US/10/098,263B
; CURRENT FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 60/276,759
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 131066
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 33827
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE: DNA
; OTHER INFORMATION: Sequence Listing Generator V 1.1
US-10-098-263B-33827

Query Match 53.3%; Score 12.8; DB 15; Length 25;
Best Local Similarity 87.5%; Pred. No. 1.8e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 AAGAAAGCAGCAGCAAG 22
   ||||| ||||| |
Db 2 AAGAAAGTACAGCATG 17
```


Mon Jul 21 15:46:54 2003

us-09-923-684-12.szlm25.rnpb

Page 5

Search completed: July 17, 2003, 21:50:23
Job time : 184 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 17:33:18 ; Search time 3896 Seconds
(without alignments)

11884.727 Million cell updates/sec

Title: US-09-923-684-2

Perfect score: 2859

Sequence: 1 actcactataggctcgagc.....gtctcaaaaaaaaaaaaaa 2859

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_estl.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: gb_gss.*

18: em_gss_hum.*

19: em_gss_inv.*

20: em_gss_pln.*

21: em_gss_vrt.*

22: em_gss_fun.*

23: em_gss_mam.*

24: em_gss_mus.*

25: em_gss_other.*

26: em_gss_pro.*

27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	386.6	13.5	553	13	BJ070365
2	327.4	11.5	410	14	BQ325425
3	276	9.7	332	13	BM144061
c	254.6	8.9	311	9	AA102407
5	235.6	8.2	300	12	BF758122
6	231.6	8.1	437	12	BF763954

7	228.4	8.0	303	14	TL1828
8	227	7.9	4125	11	BC028413
c	225.8	7.9	990	14	BM803650
10	225.2	7.9	631	17	AQ382985
11	225.2	7.9	888	12	BF037024
12	225	7.9	681	17	AQ543621
13	224.2	7.8	423	14	BM991096
c	224.2	7.8	514	9	AI754653
15	222.6	7.8	832	12	BG697217
c	222.2	7.8	891	14	BQ440084
17	221.4	7.7	379	17	AQ083206
c	220.8	7.7	327	9	AA483606
19	221	7.7	487	10	AW865946
c	220.8	7.7	640	14	BM989009
21	220.4	7.7	333	9	AA570740
c	220	7.7	416	17	AQ140695
23	219.8	7.7	544	17	AQ394650
c	219.4	7.7	455	9	AA904211
25	219.2	7.7	495	10	AW969743
c	219.2	7.7	634	14	BQ277998
27	219.2	7.7	936	17	AQ749045
c	219	7.7	537	17	AQ322719
29	219	7.7	662	17	AQ059036
30	219	7.7	686	17	AG010991
31	219	7.7	692	17	AG173021
c	219	7.7	841	17	AQ744502
33	218.8	7.7	549	17	AQ541959
34	218.8	7.7	964	12	BG483309
c	218.6	7.6	350	10	AV762430
36	218.6	7.6	448	17	AQ244402
c	218.4	7.6	471	17	AQ312810
38	218.4	7.6	504	13	BM667326
c	218	7.6	790	17	AQ740405
40	217.6	7.6	335	9	AA568204
c	217.6	7.6	360	14	BQ711799
c	217.6	7.6	507	10	AW505253
c	217.4	7.6	519	17	AQ175965
c	217.4	7.6	585	10	BE315483
44	217.4	7.6	648	17	AQ266668
45	217.4	7.6			

ALIGNMENTS

RESULT 1	BJ070365	553 bp	mRNA	linear	EST 07-DEC-2001
LOCUS	BJ070365	NIBB Mochii clone XL056f16 5', mRNA sequence.			
DEFINITION	laevis cdna	normalised Xenopus tailbud library			
ACCESSION	BJ070365				
VERSION	BJ070365.1	GI:17422350			
KEYWORDS	EST.				
SOURCE	African clawed frog.				
ORGANISM	Xenopus laevis				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopus.				
AUTHORS	Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara,Y.				
TITLE	Expressed genes in X. laevis embryo				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Tadasu Shin-I Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshin@genes.nig.ac.jp. Location/Qualifiers 1. 553 /organism="Xenopus laevis" /db_xref="taxon:8355"				
FEATURES	source				

sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-CI0156-
120101-011-g03&t3=2001-01-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 13.
Location/Qualifiers
1. 410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0156"
/dev_stage="Adult"
/note="Organ: colon.ins; Vector: puc18; Site_1: Smal;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
low stringency conditions."
BASE COUNT 93 a 143 c 93 g 81 t
ORIGIN
Query Match 11.5%; Score 327.4; DB 14; Length 410;
Best Local Similarity 96.0%; Pred. No. 8.9e-52;
Matches 389; Conservative 0; Mismatches 11; Indels 5; Gaps 5;
QY 1250 CCCCCCAGCAATACAGCTCATTCACAAATGGACAACTGGAATGCGGCCAGCTCGGAAA 1309
Db 7 CCCCCCAGGAGTATAGCTCGTT-CAATGGACAACTGGAATGCGGCCAGCTCGGAAA 65
QY 1310 CTGGAGAGCCAGTCCCC-TCGACAGCGCTGC-TGCTCTCCAGACATGCGGCCAGCTCA 1367
Db 66 CTGGAGATCCAGTCCCCCGTGCAGCGCGTGTCTCTCCAGAACTGCGGCCAGCTCA 125
QY 1368 GAAAGCAGTACCTCTGTACACGCCATCTTACAGCTGCGCTTCTCTCTACCATTTATGA 1427
Db 126 GAAAGCAGTACCTCTGTACACGCCATCTTACAGCTGCGCTTCTCTCTACCATTTATGA 185
QY 1428 CACTTCCCTCTGGACTCTCACTTCTCAGCAGCAAAAGCAATGTTGCCGGCCAAAGTTC 1487
Db 186 CACTTCCCTCTGGACTCTCACTTCTCAGCAGCAAAAGCAATGTTGCCGGCCAAAGTTC 245
QY 1488 GGGCAGCCCAAGGATCCCTTGTGAGTGGCAGCGCTTTTCTCTGAGCACAATGCCAGCC 1547
Db 246 GGGCAGCCCTTGGATCCCTTGTGAGTGGCAGCGCTTTTCTCTGAGCACAATGCCAGCC 305
QY 1548 AGCGGTGAATGCCAGTGGC-ATTATGCCAACCCCTAGTGCCTAGC-AGCTGCTCTCAG 1605
Db 306 AGCGGTGAATGCCAGTGGC-ATTATGCCAACCCCTAGTGCCTAGC-AGCTGCTCTCAG 365
QY 1606 CTAAATAATCTCCAGAGCCCGGCACTGCTAGGCACAGCC 1650
Db 366 CTAAATAATCTCCAGAGCCCGGCACTGCTAGGCACAGCC 410
RESULT 3
BM144061 332 bp mRNA linear EST 29-NOV-2001
LOCUS ADP/2.25a Differentiating glial CG-4 cells RDA library Rattus
DEFINITION norvegicus cDNA, mRNA sequence.
ACCESSION BM144061
VERSION BM144061.1 GI:17154128
/clones="XL056f16"
/clone_lib="NIBB Mochii normalized Xenopus tailbud
library"
/tissue_type="whole embryo"
/dev_stage="stage 25"
/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
BASE COUNT 136 a 133 c 139 g 144 t 1 others
ORIGIN
Query Match 13.5%; Score 386.6; DB 13; Length 553;
Best Local Similarity 83.2%; Pred. No. 6e-63;
Matches 440; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 304 GGCCTGAGCGGCTCGCAGAGAGCTGGATCGCACTTCTCGACAGCTTTGGATGAT 363
Db 25 GGACCTTGACAACGTGGCAGAGAACTGGATCTCTCTACAGACTCTTGGATGAT 84
QY 364 TTGTTTCTGTAGTACTGTATGCAAAATCATATATATCCGAGACCGCTTCTGTC 423
Db 85 TTGTTTCTGTAGTACTGTAGCGGAAATCATATATTCAGAAACTGCTCGGTC 144
QY 424 ATTAGGCTTATCCAGGTGGAGTCCAGGGCAACAGTATTTATGAATACATCCATCTT 483
Db 145 ATCTGGAGCTGTCCAGGTGGAACTGACAGGCAATAGTATATATGAATACATTCATCGT 204
QY 484 CTGACCAGCATGATGACCGTGTCTCAGCGGCCACCGAGCGCTGCACCAACCACTGC 543
Db 205 CGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 264
QY 544 TCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
Db 265 TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
QY 604 GGAACGGGCTGACCTGACGGGATACAAAGTATCCACTGCACTGCTACTTGAAGA 663
Db 325 GGAATGGGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
QY 664 TCAGGAGTATATGCTGGACATGCTCCCTGTACAGCTCTGCTACAGATTTGGGGCTG 723
Db 385 TAAGGAGTATATGCTGGACATGCTCCCTGTACAGCTCTGCTACAGATTTGGGGCTG 444
QY 724 TGGCGGTGGCCAGTCTGCCACCCAGTCCATCAGGATCAAGCTGTACAGTAACA 783
Db 445 TGGCGGTGGGTGAGTCTGCTGCCCCAGTCCATCAGGATCAAGCTGTACAGTAACA 504
QY 784 TGTTTCATGTTTCAGGCGGAGCTTGCACCTGAAGCTGATTTCTTGGATTC 832
Db 505 TGTTTCATGTTTCAGGCGGAGCTTGCACCTGAAGCTGATTTCTTGGATTC 553
RESULT 2
BQ325425 410 bp mRNA linear EST 17-MAY-2002
LOCUS PM1-CI0156-120101-011-g03 CI0156 Homo sapiens cDNA, mRNA sequence.
DEFINITION BQ325425
ACCESSION BQ325425
VERSION BQ325425.1 GI:20939694
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 410)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
TITLE

1107 GAACCTTACGCTGTCCCTGGAGCAGGTGCTCCACTGCCAAGTCCAGGACTC 1156
|||||
252 GAACCTTACGCTGTCCCTGGAGCAGGTGCTCC-CTGCCAAGTCCAGGACTC 300
|||||

RESULT 6
BF763954 437 bp mRNA linear EST 12-JAN-2001
LOCUS CM0-CS0042-301000-638-d07 CS0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF763954
VERSION BF763954.1 GI:12111854
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 437)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CM0&l2=CM0-CS0042-
301000-638-d07&t3=2008-10-30&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 437.
Location/Qualifiers
1. 437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CS0042"
/dev_stage="Adult"
/note="Organ: colon_est; Vector: puc18; Site:1: Smal;
Site:2: Smal; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 142 a 96 c 110 g 89 t
ORIGIN
Query Match 8.1%; Score 231.6; DB 12; Length 437;
Best Local Similarity 84.5%; Pred. No. 1.1e-33;
Matches 272; Conservative 0; Mismatches 49; Indels 1; Gaps 1;
QY 2538 GAAGGTGACATCAGGATTTATGTCGCCAGCGCGGCTCAGTGGCTCACACCTGTAATCCC 2597
|||||
DB 30 GAAGGTAACTATAAAATAAGTAAGTTGGCGGCGACAGTGGCTCACACCTGTAATCCC 89
|||||
QY 2598 AGCAGCTTTGGGAGGCCAAGGTGGCGGATCACCTCAGGTGAGGTTTGGCACAAGCTG 2657
|||||
DB 90 AGCAGCTTTGGGAGGCCAAGGAGGTGGATCACCTTAGTTCAGGAATTTAAGACCAAGTCTG 149
|||||
QY 2658 -CCAACTGAAACCCCTCCACTAAAAATACAAAAATAGTTGGCATGTTGGTGGT 2716
|||||
DB 150 ACCAATAGTGAACCCCTCTCTACATAAATAACAAAAATTAGCTGGGTATGTTGGTCTG 209
|||||

RESULT 5	BF758122	300 bp	mrna	linear	EST 12-JAN-2001
LOCUS	CM4-CT0574-111100-430-c04 CT0574				Human sapiens cdna, mRNA sequence.
DEFINITION	BF758122				
ACCESSION	BF758122.1	GI:12106009			
VERSION	EST.				
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 300)				
	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4st2-CM4-CT0574-111100-430-c04st3-2000-11-11st4=1) Seq primer: puc 18 forward High quality sequence start: 63 High quality sequence stop: 291. Location/Qualifiers 1..300 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT0574" /dev_stage="Adult" /note="Organ: colon; Vector: puc18; Site:1: Sma1; Site:2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 65 a 95 c 84 g 56 t				
FEATURES	source				
BASE COUNT	65 a	95 c	84 g	56 t	
ORIGIN					
Query Match	8.2%;	Score 235.6;	DB 12;	Length 300;	
Best Local Similarity	97.2%;	Pred. No. 1.9e-34;			
Matches 282;	Conservative	0;	Mismatches 4;	Indels 4;	Gaps 4;
QY	867	GACCTGATCGAAGACCCCTATACCATCATCGTGCACGGCTCGACGTTTCCACCTCCGC	926		
Db	15	GACCTGATCGAAGA-CCTATACCATCATCGTGCACGG-TCCGACGTTG-TACACCTCCGC	71		
QY	927	TACGCACACCACTCTGTTGGTGAAGGGCCAGCTACCAACCAAGTACTACCGGCTGCTG	986		
Db	72	TACGCACACCACTCTGTTGGTGAAGGGCCAGGTCACCAACCAAGTACTACCGGCTGCTG	131		
QY	987	TCCACAGCGGGGGCTGGGTGTGGTGTGCAGAGCTACGCCACCGTGGTGCACACAGCCGC	1046		
Db	132	TCCACAGCGGGGGCTGGGTGTGGTGTGCAGAGCTACGCCACCGTGGTGCACACAGCCGC	191		
QY	1047	TGCTCCGCCGCCACTGCATCGTGAAGTTCATATGCTACTACCGAGATTGAATACAAG	1106		
Db	192	TGCTCCGCCGCCACTGCATCGTGAAGTTCATATGCTACTACCGAGATTGAATACAAG	251		

QY	2717	AGCACCCTGTATCCAGCTACTCTGGAGGCTGAGATAGGAGGATCACTGAACCCGGGAG	2776
Db	210	CACCCCTGTATCCAGCTACTCTAGGAGGCTGAGGCAGGAATCACTTGAACCTGGAAG	269
QY	2777	GTGAGGTTGCAGTGAGCTAAGATCACATCACTCCACTCCAGCTCCGGTAAACAGAGTGAG	2836
Db	270	GCAGAGTTGCAGTGAGCCAGGATCAACCACCTGCACCCCTAGCCTGGCGGCACAGACGAAG	329
QY	2837	ACTGCTCTCAAAAAAAAAAAAAA	2858
Db	330	ACTGCTCTCAAAAAAAAAAAAAA	351
RESULT	7		
LOCUS	T11828		
DEFINITION	T11828	303 bp mRNA linear EST 28-NOV-1994	
ACCESSION	A1049f	Heart Homo sapiens cDNA clone A1049 similar to Alu	
VERSION		repetitive element, mRNA sequence.	
KEYWORDS	T11828.1	GI:596532	
SOURCE	EST.	human.	
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 303)		
JOURNAL	Liew,C.C., Hwang,D.M., Fung,Y.W., Laurensen,C., Cukerman,E., Tsui		
MEDLINE	S. and Lee,C.Y.		
COMMENT	A catalogue of genes in the cardiovascular system as identified by		
	expressed sequence tags		
	Proc. Natl. Acad. Sci. U.S.A. 91, 10645-10649 (1994)		
	95034171		
	Other ESTs: A1049R		

```

seq primer: GGTGGCGACACTCTCTGGAGCC.
Location/Qualifiers
1..303
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="A1049"
  /clone_1lb="Heart"
  /lab_host="E.coli Y1090"
  /notes="vector: Lambda gtl1; Site_1: EcoRI; Site_2: EcoRI"
BASE COUNT      93 a    71 c
ORIGIN
Query Match      8.0%; Score 228.4; DB 14; Length 303;
Best Local Similarity 86.1%; Pred. No. 4.3e-33;
Matches 253; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
Qy 2566 GGCGGGGCTCAGTGGCTCACACCTGTAAATCCCGACACTTTGGGAGGCCCAAGTGGGGGA 2625
      |||||
Db 2 GGCGGGGCGACAGTGGCTCACACCTGTAAATCCCGACACTTTGGGAGGCCCAAGTGGGTGA 61
      |||||
Qy 2626 TCACCTTGAGGTTCAGAGTTTTCGACAAAGCTGCCAACAAAGCTCAAAACCCCATCTCCACATA 2685
      |||||
Db 62 TCACCTTGAGGTTCAGAGTTTTCGACAAAGCTGCCAACATGGTAAAGCCCATCTCTACTA 121
      |||||
Qy 2686 AAATAACAAAATTAAGTTGGGATGTGTGGGATGTGTGGACCACTGTAAATCCAGCTACTCTGGAGG 2745
      |||||
Db 122 AAATAACAAAATTAAGTTGGGATGTGTGGGATGTGTGGACCACTGTAAATCCAGCTACTCTGGAGG 181
      |||||
Qy 2746 CTGAGATAGGAGGATCACTTTGAACCCGGGAGTTGGAGTTTGGAGTGAGCTTAAGATCAAT 2805
      |||||
Db 182 CTGAGGCGAGGAATCACATTGACTCTGTGAGGCGAGGTTGCAGTGAGCTGAGTTGGCAC 241
      |||||
Qy 2806 CACTGCACCTCCAGCCTGGGTACAGAGTGAGACTGTCTCAAAAAA
      |||||

```

Db	242	CACTGCATCCAGCCTGGCAACAGAGAGACTCCATCTCAAAAAAAAAAAAA	295
RESULT 8	BC028413		
LOCUS	BC028413		
DEFINITION	Homo sapiens, clone IMAGE:4823388, mRNA.		
ACCESSION	BC028413		
VERSION	BC028413.1	GI:22382098	
KEYWORDS	HTC.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 4125)		
JOURNAL	Strausberg, R.		
	Direct Submission		
	Submitted (23-APR-2002) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		
	CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki		
	Toshiyuki and Piero Carninci (RIKEN)		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Institute for Systems Biology		
	http://www.systemsbio.org		
	contact: amadan@systemsbiology.org		
	Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha		
	Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting		
	Clone distribution: MGC clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Series: IRAC Plate: 46 Row: k Column: 18		
	This clone was selected for full length sequencing because it		
	passed the following selection criteria: Hexamer frequency ORF		
	analysis		
	This clone has the following problem: incomplete processing.		

FEATURES	this clone has the following problem: incomplete processing.									
source	Location/Qualifiers									
	1. 4125									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	/clone="IMAGE:4823388"									
	/tissue_type="Testis"									
	/clone_lib="NIH_MGC_97"									
	/lab_host="DH10B"									
	/note="Vector: pBluescript"									
BASE COUNT	985	a	1109	c	1185	g	846	t		
ORIGIN										
Query Match	7.9%; Score 227; DB 11; Length 4125;									
Best Local Similarity	87.8%; Pred. No. 7.8e-33;									
Matches	259;	Conservative	0;	Mismatches	35;	Indels	1;	Gaps	1;	
Qy	2566	GGCGGGGCTCAGTGGGTACACCTGTAATCCGAGCACTTTGGGAGGCCAAGGTGGCGGA	2625							
Db	3552	GGCAGCGCGAGTGGCTCATCTGTATCCGAGCACTTTGGGAGGCCAAGCGCGCGGA	3611							
Qy	2626	TCACCTGAGTCCAGGAGTTGGGACAGCGCTG-CCAACACAGCTGAACCCCATCTCCACT	2684							
Db	3612	TCACTTGGCGTCCAGGAGTTCGAGACAGCGCTGACCAACATGGAGAAACCCCATCTCCACT	3671							
Qy	2685	AAATAACAAAATTAGTTGGGCATGGTGGTGAGCACCTGTAAATCCGAGTACTCTCGGAG	2744							
Db	3672	AAAGTACAAAAATTAGCTGGGCGGTGGTGGCGCATCTGTAAATCCGAGTACTCGGAG	3731							
Qy	2745	GCTGAGATAGGAGGATCACTTGAACCCGGGAGGTGGAGGTTGCGAGTGAGCTAAATCACA	2804							
Db	3732	GCTGAAGCAGGAGATATCACTTGAACCCAGGAGGAGGTTGCGAGTGAGTCAAGTCAACG	3791							

QY	2805	TCATGCACTCCAGCCTGGGTAAACAGAGTGAGACTGTCTCAAAAAAAAAAAAAA	2859
Db	3792	CCATGTCACCTCCAGCTGGGCAACAGAGTGAGACTGTCTCAAAAAAAAAAAAAA	3846
RESULT 9			
BM803650/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
BM803650 990 bp mRNA linear EST 05-MAR-2002			
AGENCOUT_6439541 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5520297			
5', mRNA sequence.			
BM803650			
BM803650.1	GI:19120473		
EST.			
human.			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 990)			
NIH-MGC	http://mgc.nci.nih.gov/.		
National Institutes of Health, Mammalian Gene Collection (MGC)			
Unpublished (1999)			
Contact: Robert Strausberg, Ph.D.			
Email: cgapbs.r@mail.nih.gov			
Tissue Procurement: ARCC			
cDNA Library Preparation: Life Technologies, Inc.			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Agencourt Bioscience Corporation			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: LLAM12184	row: b column: 10		
High quality sequence stop: 632.			
Location/Qualifiers			
1..990			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="IMAGE:5520297"			
/clone_lib="NIH_MGC_71"			
/tissue_type="leiomyosarcoma"			
/lab_host="DH10B (phage-resistant)"			
/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: NotI;			
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.			
Average insert size 2.1 kb."			
BASE COUNT	170 a. 277 c 268 g 275 t		
ORIGIN			
Query Match	7.98; Score 226.8; DB 14; Length 990;		
Best Local Similarity	83.5%; Pred. No. 8.5e-33;		
Matches 269; Conservative	0; Mismatches 52; Indels 1; Gaps 1;		
QY	2539	AAGGTGACATCAGGATTATGTGCCAGCGGGCTCAGTGGCTACACCTGTAATCCCA	2598
Db	353	AAGATTCTTAAATAATTAACCTTGGCTGGCGGGTGGCTATGCTGTAATCCCA	294
QY	2599	GCATTTGGGAGGCCAAGGTGGGCGGATACCTGAGTCAGGAGTTTGCACAGCCT-G	2657
Db	293	GCACCTTTAGGAGCGCGAGGGGGTGGATACCTGAGTCAGGAGTTTGCAGACAGCCTG	234
QY	2658	CCACAAGCTGAACCCCATCTCCACTAAAAATAACAAAAATTAGTTGGGATGGTGGTGA	2717
Db	233	CCACATGCTGAACCTTGCTCTCTACTAAAAACACAAAAATTAGCCTGGCATGGTGGCGG	174
QY	2718	GCACCTGTAATCCAGCTACTCTCGGAGCTGAGATAGAGGATCATTGTAACCCGGGAGG	2777
Db	173	GGCCCTGTAATCCAGCTACTCTCGGAGGCTGAGGAGGATCACATGTGATCCTGGGAGG	114
QY	2778	TGGAGTTGCACTGAGCTAAGATCACATCTGCCTCCAGCCTGGGTGAACAGAGTGA	2837
Db	113	TGGAGTTGCACTAAGCCACCATCACACCTGCACCTCCAGCCTGGGTGACAGAGAGA	54
QY	2838	CTGTCTCAAAAAAAAAAAAAA	2859

QY	2679	TCACATAAAATACAAAAATAGTTGGCATGGTGGTGGACACCTGTAATCCAGCTACT	2738
Db	138	TCTACTAAAAACACAAAAATAGCTGTGTGTGGTGGTGCACACCTATAATCCAGCTACT	197
QY	2739	CTGAGGCTGAGTAGGAGGATCACTTGAACCGGAGGTGGAGGTGGCAGTGCAGCTAAG	2798
Db	198	TGGGAGGCTAAGCGAGGAGATCACTTGAACCTGGGAGCGGAGGTGGCAGTGCAGCTAG	257
QY	2799	ATCACATCACTGCACCTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAAAAAA	2858
Db	258	ATCGCACCACTGCACCTCCAGCTGGGCCACAGAGTGAGACTGTCTCTCAAAAAA	317
QY	2859	A	2859
Db	318	A	318
RESULT 13			
BM991096			
LOCUS			
DEFINITION	423 bp mRNA linear EST 17-JUN-2002		
ACCESSION	BM991096		
VERSION	BM991096.1 GI:19710485		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-re@mail.nih.gov		
	Tissue Procurement: Dr. Jose Mercuende		
	CDNA Library preparation: Dr. M. Bento Soares, University of Iowa		
	CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: Clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	The following repetitive elements were found in this cDNA		
	sequence: 1-21. >AtRichLow_complexity 77-366, >ALU		
	Seq primer: M13 FORWARD		
	POLYA-Yes.		
FEATURES			
source	Location/Qualifiers		
	1. .423		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5862692"		
	/clone_lib="NCI_CGAP_D10"		
	/tissue_type="Lung Focal Fibrosis"		
	/dev_stage="Adult"		
	/lab_host="DHI08 (Life Technologies)"		
	/note="Organ: Lung; Vector: pF73-Pac (Pharmacia) with a		
	modified polylinker; Site_1: EcoR I; Site_2: Not I;		
	NCI_CGAP_D10 is a cDNA library containing the following		
	tissue(s): A pool of Lung Focal Fibrosis. The library was		
	constructed according to Bonaldo, Lennon and Soares,		
	Genome Research, 6:791-806, 1996. First strand cDNA		
	synthesis was primed with an oligo-dT primer containing a		
	Not I site. Double stranded cDNA was ligated to an EcoR I		
	adaptor, digested with Not I, and cloned directionally		
	into pF73-Pac vector. The oligonucleotide used to prime		
	the synthesis of first-strand cDNA contains a library tag		
	sequence that is located between the Not I site and the		
	(d)18 tail. The sequence tag for this library is		
	ATACCGCGTC.		
	TAG_LIB=UI-H-D10		
	TAG_TISSUE=lung with fibrosis		
	TAG_SEQ=ATACCGGTC"		
BASE COUNT	111 a 100 c 111 g 101 t		
ORIGIN			
	Query Match 7.8%; Score 224.2; DB 14; Length 423;		
	Best Local Similarity 84.3%; Pred No. 2.6e-32;		
	Matches 264; Conservative 0; Mismatches 48; Indels 1; Gaps 1;		
QY	2548	TCAGGATTATGTGCCCCAGCGCGGCTCAGTGGCTCACACCTGTAATCCAGCACTTTGG	2607
Db	59	TCAAAATTAACCTGACTTGGCTGGCGCAGTGGCTCACACCTGTAATCCAGCACTTTGG	118
QY	2608	GAGGCCAAGGTGGCGGATCACTGAGGTTCAGAGTTTTCGACAAAGCCTG-CCAACAAGC	2666
Db	119	GAGGCCAAGGGAGGGGATCACTGAGTTCAGAGTTTCGAGACCGCTGATCGCATGG	178
QY	2667	TGAACCCCATCTCCCACTAAAAATACAAAAATAGTTGGGCATGGTGGTGCACCTGTA	2726
Db	179	TGAACCCCGTTTCTACTAAAAATACAAAAATAGCCGGCGCTGGTGGTGCCTGTGA	238
QY	2727	ATCCCAAGTACTCTGGAGCTGAGATAGGAGGATCACTTGAACCGGAGGTGGAGTTG	2786
Db	239	ATCCCAAGTACTCAGGAGGCTGAGGAGGAGATCACTTGAACCGGAGGTGGAGTTG	298
QY	2787	CAGTGAGCTAAGATCACATCACTGCACCTCGAGCTGGGTAAACAGAGTGAGACTGTCTCA	2846
Db	299	CAGTGAGCTGAGATCGTGCCTGCCTGCAGCTGGCGCACAGAGCAAACTCCGCTC	358
QY	2847	AAAAAAAAAAAAA	2859
Db	359	AAAAAAAAAAAAA	371
RESULT 14			
AI754653/c			
LOCUS			
DEFINITION	514 bp mRNA linear EST 20-JUN-2002		
ACCESSION	AI754653		
VERSION	AI754653.1 GI:5132917		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	Jia, L., Young, M.F., Powell, J., Yang, L., Ho, N.C., Hotchkiss, R.,		
MEDLINE	Robey, P.G. and Francomano, C.A.		
COMMENT	Gene expression profile of human bone marrow stromal cells: high-throughput expressed sequence tag sequencing analysis		
	Genomics 79 (1), 7-17 (2002)		
	Contact: Libin Jia		
	Medical Genetics Branch		
	National Human Genome Research Institute		
	10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA		
	Tel: 301-402-4877		
	Fax: 301-496-7157		
	Email: libin@helix.nih.gov		
	DNA Sequencing and analyses by National Institutes of Health		
	Intramural Sequencing Center (NISC).		
	Plate: 28 row: C column: 08		
	Seq primer: -21M13 forward primer (ABI).		
	Location/Qualifiers		
	1. .514		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="HBMSC_cr28c08"		
	/clone_lib="Human bone marrow stromal cells"		
	/sex="mixed"		
	/tissue_type="bone marrow stroma"		
	/dev_stage="mixed"		
	/lab_host="XLI-Blue MRF//SOLR"		
	/note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;		
	mRNA made from human bone marrow stroma, cDNA made by		

oligo-dT priming. Directionally cloned. Size-selected for average insert size >0.5 kb. Library constructed by Dr. Marian Young and Dr. Pamela Gehron Robey (NIDCR). Library supplied by Dr. Libin Jia (NIGRI)"

[illegible]

```

RESULT 15
BG697217
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 832)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAMI0699 row: f column: 14
High quality sequence stop: 694.
Location/Qualifiers
1. 832
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:4803469
/clone_lib="NCI_CGAP_Skn3"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skln; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5Kb. Library constructed by Life

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 17:35:07 ; Search time 131 Seconds
(without alignments)
6693.045 Million cell updates/sec

Title: US-09-923-684-2
Perfect score: 2859
Sequence: 1 actcactataggctcgagc.....gtctcaaaaaaaaaaaaaa 2859

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	220.6	7.7	70000	4	US-09-851-896-3
C 2	219.8	7.7	15977	4	US-09-608-285A-59
C 3	219.6	7.7	72928	3	US-09-009-913-1
C 4	218.2	7.6	53526	3	US-08-658-136-2
C 5	218.2	7.6	53577	3	US-08-658-136-1
C 6	217.8	7.6	36741	4	US-09-301-665-3
C 7	217.8	7.6	11282	4	US-09-754-250-3
C 8	217	7.6	8758	4	US-09-799-345-3
C 9	216.6	7.6	152331	3	US-09-128-155-16
C 10	215.8	7.5	80246	4	US-09-078-294-3
C 11	215.8	7.5	80595	4	US-09-167-681-45
C 12	215.2	7.5	8453	4	US-09-729-995-3
C 13	215.2	7.5	29629	4	US-09-729-995-3
C 14	213.8	7.5	2532	4	US-09-799-345-1
C 15	212.6	7.4	11531	1	US-08-068-945A-1
C 16	212.6	7.4	11531	1	US-08-442-806-1
C 17	212.6	7.4	21234	4	US-09-810-671-3
C 18	212	7.4	36741	4	US-09-301-665-3
C 19	211.6	7.4	282	1	US-08-133-628-8
C 20	211	7.4	99500	4	US-09-798-096-10
C 21	211	7.4	168575	4	US-09-426-290-1
C 22	210.6	7.4	55827	4	US-09-813-133A-3
C 23	210.4	7.4	3867	4	US-09-347-114A-81
C 24	210.4	7.4	14581	4	US-08-520-373D-4
C 25	210.4	7.4	22481	4	US-08-367-841A-43
C 26	210.4	7.4	22481	5	PCT-US95-07201-43
C 27	210.4	7.4	22484	4	US-09-875-223-2

C 28	210.4	7.4	43950	4	US-09-735-934A-3	Sequence 3, Appli
C 29	210.2	7.4	15977	4	US-09-608-285A-59	Sequence 59, Appl
C 30	209.6	7.3	14747	4	US-09-608-285A-42	Sequence 42, Appl
C 31	209.6	7.3	80246	4	US-09-078-294-4	Sequence 4, Appli
C 32	209.6	7.3	80595	4	US-09-078-294-3	Sequence 3, Appli
C 33	209.2	7.3	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 34	209.2	7.3	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 35	209.2	7.3	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 36	208.6	7.3	31571	1	US-08-323-443B-1	Sequence 1, Appli
C 37	208.6	7.3	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 38	208.6	7.3	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 39	208.6	7.3	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 40	208	7.3	81001	4	US-09-750-580-1	Sequence 1, Appli
C 41	207.8	7.3	2115	1	US-08-395-800A-7	Sequence 7, Appli
C 42	207.8	7.3	4823	2	US-08-457-254-5	Sequence 5, Appli
C 43	207.8	7.3	4823	2	US-08-484-257-20	Sequence 20, Appl
C 44	207.8	7.3	4823	3	US-08-999-927-5	Sequence 5, Appli
C 45	207.8	7.3	4823	4	US-08-461-819-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-851-896-3/c
; Sequence 3, Application US/09851896
; Patent No. 6410325

; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett

; APPLICANT: Susan M. Freier

; APPLICANT: Andrew T. Watt

; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEP

; TITLE OF INVENTION: EXPRESSION

; FILE REFERENCE: RTS-0220

; CURRENT APPLICATION NUMBER: US/09/851,896

; CURRENT FILING DATE: 2001-05-08

; NUMBER OF SEQ ID NOS: 89

; SEQ ID NO 3

; LENGTH: 70000

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

US-09-851-896-3

Query Match 7.7%; Score 220.6; DB 4; Length 70000;
Best Local Similarity 86.4%; Pred. No. 4.5e-44;
Matches 255; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY	2563	CCAGCGCGGCTCAGTGGCTCACACCTGTATCCAGCACCTTTGGGAGGCCAAGGTGGGC	2622
Db	12155	CTAGGCCAGGTGCAGTGGCTCACCGCTGCAATCCAGCACCTTTGGGAGGCCAAGGTGGGC	12096
QY	2623	GGATCAGCTGAGTTCAGAGTGTTCGCAACAGCCT-GCCAAACAGCTGAACCCCATCTCC	2681
Db	12095	AGATCAGCTGAGTTCAGAGTGTTCGCAACAGCCTGGCCAAATGATGAACCCCATCTCT	12036
QY	2682	ACTAAATACAAAATTTAGTGGCATGTGGTGGAGCAGCTGTATCCAGTACTCTG	2741
Db	12035	GCTAAATACAAAATTTAGTGGCATGTGGTGGAGCAGCTGTATCCAGTACTCTG	11976
QY	2742	GAGCTGAGATAGGAGTATCATTTGAACCGGAGGTGGAGTTCAGTGAAGTATGATC	2801
Db	11975	GAGCTGAGGAGGAGTATCATTTGAACCGGAGGTGGAGTTCAGTGAAGTATGATC	11916
QY	2802	ACATCACTGAGTTCAGGCTGGGTACAGAGTGAAGTCTCTCAAAAAA	2856
Db	11915	GCCTCACTGAGTTCAGGCTGGGTACAGAGTGAAGTCTCTCAAAAAA	11861

RESULT 2

US-09-608-285A-59

; Sequence 59, Application US/09608285A

; Patent No. 6335013

QY 2619 GGGGGATCCTCAGGTGAGGCTTGGGACAGGCTT-CCCAACAGCTGAACCCCAT 2677
|||||
Db 9813 GGGGGATCCTCAGGTGAGGCTTGAAGACAGGCTGGCCAAACATGATGAACCCCAT 9754
|||||
QY 2678 CTCACATAAAATAACAAAATTAGTTGGGCTAGGTGAGGCTGTAATCCAGCTAC 2737
|||||
Db 9753 CTCCTACTAAAAATGCAAAAATTAGCCAGGCTGTTGGCATGCCACCTCTGGTCCAGCTAC 9694
|||||
QY 2738 TCTGGAGGCTGAGATAGGAGGATCCTTGAACCCGGAGGTGGAGGTTGCAGTGAGCTAA 2797
|||||
Db 9693 TTGGGAGGCTGAGCGCAGGAGAAATCATTTGAACCCAGTAGCGGAGGTTCCAGTGAGCTGA 9634
|||||
QY 2798 GATCACAATCCTGACTCCAGCTGGGTAACAGAGTGACATGTTCTCAAAAAA 2857
||
Db 9633 AATTGTGCCACTGCACTCCAGCTGGGCAACAGAGTGAGACTCTGTCAAAAAA 9574
||
QY 2858 AA 2859
||
Db 9573 AA 9572

RESULT 4

US-08-658-136-2/c
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53526 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match 7.6%; Score 218.2; DB 3; Length 53526;
Best Local Similarity 81.7%; Pred. No. 1.6e-43;
Matches 264; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 2538 GAAGGTGACATCAGGATTATGTCGCCAGCGGCTCAGTGGCTCAGACCTGTAATCCC 2597

Db 1314 GAAGGAGAAAAAACAAGAACCCAGGCCACAGTGCTCACACCTGTATATCCC 1255
|||||
QY 2598 AGCACTTTGGGAGGCCAAGGTGGCGGATCACCTGAGGTGAGGCTTGGGACAAACCTT- 2656
|||||
Db 1254 AGCACTTTGGGAGGCCAAGCGCGGGTGCATGAGGCCAAGAGTTCGAGACCAACCTG 1195
|||||
QY 2657 GCCAACAAGCTGAACCCCATCTCCACTAAAAATACAAAAATAGTTGGGCTGTTGGTG 2716
|||||
Db 1194 GCCAAGCTGCAAAACCCCGTCTCTACTAAAAATACAAAAATAGTTGGGCTGTTGGTG 1135
|||||
QY 2717 AGCACTGTATATCCAGCTACTCTGGAGGCTGAGATAGGAGGATCACTTGAACCCGGAG 2776
|||||
Db 1134 GGTGCTGTATATCCAGCTACTCTGGAGGCTGAGCGAGAGATCACTTGAACCCGGAG 1075
|||||
QY 2777 GTGGAGTTCAGTGAGCTAAGATCATCATCTCCAGCTGCTCCAGCTGGGTGAGAGTGA 2836
|||||
Db 1074 GCGGAGGTGCGAGTGAGCTGAGGTTACCCCACTGCACTCCAGCTGGGTGAGAGTGA 1015
|||||
QY 2837 ACTGTCTCAAAAAA 2859
|||
Db 1014 ACTCCATCTCAAAAAA 992

RESULT 5

US-08-658-136-1/c
; Sequence 1, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

Query Match 7.6%; Score 218.2; DB 3; Length 53577;
Best Local Similarity 81.7%; Pred. No. 1.6e-43;
Matches 264; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

10660 AGTCCTGGCGACAGAGACTCTATCTCAAAAGAAAAA 19608

[illegible]

RESULT 6
US-09-301-665-3/c
; Sequence 3, Application US/09301665
; Patent No. 6207876
; GENERAL INFORMATION:
; APPLICANT: KELLEMS, RODNEY E.
; APPLICANT: DATTA, SURJIT K.
; APPLICANT: BLACKBURN, MICHAEL R.
; TITLE OF INVENTION: ADENOSINE DEAMINASE DEFICIENT TRANSGENIC MICE AND
; TITLE OF INVENTION: METHODS FOR THE USE THEREOF
; FILE REFERENCE: UTSH:243
; CURRENT APPLICATION NUMBER: US/09/301,665
; CURRENT FILING DATE: 1999-04-28
; EARLIER APPLICATION NUMBER: 60/083,408
; EARLIER FILING DATE: 1998-04-29
; EARLIER APPLICATION NUMBER: 60/083,370
; EARLIER FILING DATE: 1998-04-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 36741
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-301-665-3

Query Match	7.6%	Score	217.8	DB 4	Length	36741			
Best Local Similarity	84.0%	Pred. No.	1.7e-43						
Matches	246	Conservative	0	Mismatches	47	Indels	0	Gaps	0
QY	2567	GCCGGGCTCAGGGCTCACACTGTAATCCAGACACTTTGGAGGCCAAGGTGGCGGAT	2626						
Db	19900	GCCAGCACAGTGGCTCATGCTGTAATCCAGACACTTTGGAGGCCAAGGTGGGTGGAT	19841						
QY	2627	CACCTGAGGTCAGAGGAGTTTGGACAAAGCCTGCCAACAGCTGAAACCCCACTCTCCACTAA	2686						
Db	19840	CACCTGAGGTCAGGAGTTCGAGCACCAGCCTGGCCAAACATGTGAAACCCCACTCTCCGCTTAA	19781						
QY	2687	AAATACAAAANTAGTTGGGCATGTTGGTGAGCACCTGTAATCCAGACTACTCTGGAGGC	2746						
Db	19780	AAATACAAAANTTAGCTGATGTGGTGGCATGTGCTTGTATCCAGACTACTCTAGGAGGC	19721						
QY	2747	TGAGATAGGAGGATCACTTGAACCCGGAGGTGGAGTTGCAGTCAGTAAAGTACACATC	2806						
Db	19720	TGAGGCAGGAGAAATCACTTGAACCCAGGAGGTGGAGTTGCACTGAGCCGAGATCATGCC	19661						
QY	2807	ACTGCACTCCAGCGCTGGGTAAACAGCTGAGACTGTCTCAAAAAAATAAAAAA	2859						

Db

[illegible]

Query Match	7.6%;	Score 217.8;	DB 4;	Length 111282;
Best Local Similarity	85.5%;	Pred. No. 2.6e-43;		
Matches	254;	Mismatches 42;	Indels 1;	Gaps 1;
	Conservative	0;		
QY	2564	GAGGCGGGCTCAGTGGCTCACACCTGTAATCCAGACACTTTGGAGAGCCAAAGTGGGCG	2623	
Db	18618	GAGGCGAGTGTGATGGCTCAGGCGCTGTAATCCAGCAGCTTTGGGAGGTCAAGCGGGCA	18559	
QY	2624	GATCACCTGAGTTCAGGAGTTTGCACAGCCT-CCCAACAGCTGAAACCCCATCTCCA	2682	
Db	18558	GATCACCTGAGGCGAGGATTCGAGACAGACCTGCCCATAATGTTGAACCCCGTCTCTA	18499	
QY	2683	CTAAAAATACAAAAATTAGTTTGGGCATGTTGTGAGCAGCTGTATCCAGCTACTCTGG	2742	
Db	18498	CCAAAAAGTACAAAAATTACCGGGCATGTGTCAGGTGCGCTGTATCCAGCTACTCTGG	18439	
QY	2743	AGGCTCAGATAGAGGATCACCTTGAACCCGGGAGGTGGAGTTCAGCTGAGCTAAGATCA	2802	
Db	18438	AGGCTGAGCGAGCAACTCACCTGTAACCCAGGAGTGGAGTTCAGCTGAGACAAGATCG	18379	
QY	2803	CATCACTGCACCTCCAGCCTGGGTACAGAGTGCAGCTGCTCAAAAAAAAAAAAAA	2859	
Db	18378	CAGCACTGCACCTCCAGCCTGGGCGAGCAGACAGTCTCTCAAAAAAAAAAAAAA	18322	

```

RESULT 8
US-09/799-345-3/c
; Sequence 3, Application US/09799345
; Patent No. 6323016
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001156
; CURRENT APPLICATION NUMBER: US/09/799,345
; CURRENT FILING DATE: 2001-03-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 8758
; TYPE: DNA
; ORGANISM: Human
US-09/799-345-3
Query Match          7.6%; Score 217; DB 4; Length 8758;

```

Best Local Similarity 81.2%; Pred. No. 1.6e-43;
Matches 264; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
QY 2536 GTGAAGTGCATCAGGATATGTGCCAGGCGGGCTCAGTGGCTCACCTGTGAATC 2595
DB 6356 GTGAATAGAGTAAGAATGACTTCCTTTAGGCCAGGACGGTGGCTCAGCCTATATC 6297
QY 2596 CCAGCACTTTGGGAGGCCAAGGTGGCGGATCACTCAGGTGAGGATTTGGCACAAGCC 2655
DB 6296 CCAGCACTCTGGAAGCCCTAGGTAGGTGATCACTTGGAGTCTGAGGTTTGAGACCGCC 6237
QY 2656 T-GCCACAAGCTGAACCCCATCTCCACTTAAATAATACAAAAATAGTTGGCATGGTG 2714
DB 6236 TGCCCAACATGTGTGAACCCCATCTCTACTTAAATAATACAAAAATAGTGGCATGGTG 6177
QY 2715 TGAGCACTGTATATCCAGCTACTCTGGAGGCTGAGATAGGAGGATCACTTGAACCCGG 2774
DB 6176 TGACCCCTATATCCAGCTACTCTAGGAGGCTGAGGAGGAGAAATGCTTGAACCCAGG 6117
QY 2775 AGTGGAGTTCAGTGTAGTAAAGATCACTACATCACTGCAGCTCCAGCTGGTAAACAGATG 2834
DB 6116 AGTGGAGTTCAGTGTAGTAAAGATCACTACATCACTGCAGCTCCAGCTGGTAAACAGATG 2834
QY 2835 AGACTGTCTCAAAAAAATAAAAAA 2859
DB 6056 AGAGAGACTCTCTATTATTAATAAAAA 6032

RESULT 9
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER FILING DATE: 1998-07-02
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 7.6%; Score 216.6; DB 3; Length 152331;
Best Local Similarity 84.0%; Pred. No. 5.7e-43;
Matches 268; Conservative 0; Mismatches 49; Indels 2; Gaps 2;
QY 2542 GTGACATCAGGATATGTGCCAGGCGGGCTCAGTGGCTCACCTGTGAATCCAGCA 2601
DB 150967 GTGAAGTGAAGATTAATCTATCCAGGCCAGGACAGTGGCCATGCTGTATCCAGCA 151026
QY 2602 CTTTGGAGGCCAAGTGGCGGATCACTGAGGTGAGGATTTGGCACAAGCCTG-CCA 2660
DB 151027 CTTTGGAGGCCAAGTGGCGGATCACTGAGGTGAGGATTTGAAGACCAACCTGTATCA 151086
QY 2661 ACAAGTGAACCCCATCTCCACTTAAATAATACAAAAATAGTTGGCATGGTGGTGACGA 2720
DB 151087 ACATGTGAACCCCATCTCTACTTAAATATC-AAAATAGTGGCATGGTGGTGACGA 151145
QY 2721 CCTGTAAATCCAGCTACTCTGAGGCTGAGATAGGAGGATCACTTGAACCCGGAGGTGG 2780

DB 151146 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAATCAATTTGAACCTGGGAGGTGG 151205
QY 2781 AGTTGTCAGTGCAGTAAGATCACAATCACTCACTCCAGCTGGGTAAACAGAGTGAGACTG 2840
DB 151206 AGTTGTCAGTGCAGTAAGATCAGTGGTCCACTGCACTCCAGCTGGGTGACAGGAGACTCCG 151265
QY 2841 TCTCAAAAAAATAAAAAA 2859
DB 151266 TCTCAAAAAAATAAAAAA 151284
RESULT 10
US-09-078-294-4/c
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contlg
US-09-078-294-4

Query Match 7.5%; Score 215.8; DB 4; Length 80246;
Best Local Similarity 85.4%; Pred. No. 7.1e-43;
Matches 252; Conservative 0; Mismatches 42; Indels 1; Gaps 1;
QY 2561 CCCAGGCGGGCTCAGTGGCTCACACCTGTAAATCCAGCAAGCTTGGGAGGCCAAGTGG 2620
DB 14331 CTCGGGCCAGGCGCTGCTGCTTACGCTGTAAATCCAGCAGTTTAGGAGGTGCGAGTGG 14272
QY 2621 GCGGATCACTCAGTGCAGGAGTTTGGCACAAGCTT-GCCACAAGCTGAAACCCCATCT 2679
DB 14271 GCGGATCACTCAGTGCAGGAGTTTGGCACAAGCTTGGCACAAGCTGAAACCCCATCT 14212
QY 2680 CCACATAAATAACAAAAATAGTTGGGCGATGGTGGTGGCAGCTGTATCCAGCTACTC 2739
DB 14211 CTACAAAAAACACAAAAATAGTGGCGCATGGTGGCGCTATAGTCCAGCTACTC 14152
QY 2740 TGGAGGCTGAGATAGGAGGATCACTTGAACCCGGAGGTGGAGTTGCAGTGAGCTAAGA 2799
DB 14151 AGGAGGCTGAGGATGAGATCGCTTGAACCCGGGAGGAGTTGCAGTGAGCTGTGA 14092
QY 2800 TCACATCACTCACTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAAAAAAATAAAAA 2854
DB 14091 TCAGTCACTCACTCCAGCTCCAGCTAGGCAACAGTGAGACTCGATCTCAAAAAAATAAAAA 14037

RESULT 11
US-09-078-294-3/c
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA

```
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3
Query Match          7.5%; Score 215.8; DB 4; Length 80595;
Best Local Similarity 85.4%; Pred. No. 7.1e-43;
Matches 252; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 2561 CCCAGCGCGGCTCAGTGGCTCACACCTGTAATCCAGCACTTTGGGAGGCCAAGTGG 2620
Db 14598 CTCTGGCCAGGCATGGTGGCTTACGCCCTGTAATCCAGCAGTTTAGGAGGTCGAGTGG 14539

QY 2621 GCGGATCACTGAGTGCAGGAGTTTGCACACGCT-CCCAACAGCTGAAACCCCATCT 2679
Db 14538 GCGGATCACTGAGTGCAGGAGTTTGCACACGCTGCGCCACATGTTGTAACCCCATCT 14479

QY 2680 CCACATAAAATACAAAAATTAGTTGGCATGGTGTGAGCACCTGTATATCCAGCTACTC 2739
Db 14478 CTACAAAAACACAAAAATTAGTTGGCATGGTGTGAGCACCTGTATATCCAGCTACTC 14419

QY 2740 TGGAGGCTGAGATGAGGATCATTGAAACCGGAGGTGGAGTTGCAGTGAGTAAGA 2799
Db 14418 AGGAGGCTGAGCATGAGAACTGCTTGAACCGGAGGCGGAGGTTGCAGTGAGCTGTA 14359

QY 2800 TCACATCACTGCACCTCCAGCCTGGGTAACAGAGTGAGACTGTCTCAAAAAA 2854
Db 14358 TCAGTCACTGCATCCAGCCTAGGCAACAGTGAGACTGCATCTCAAAAAA 14304

RESULT 12
US-09-167-681-45/c
; Sequence 45, Application US/09167681A
; Patent No. 6265561
; GENERAL INFORMATION:
; APPLICANT: Weinshilboum, M.D., Richard M.
; APPLICANT: Raftogiannis, Rebecca B.
; APPLICANT: Wood, Thomas C.
; APPLICANT: O'therness, Diane M.
; TITLE OF INVENTION: SULFOTRANSFERASE SEQUENCE VARIANTS
; FILE REFERENCE: 07039/118001
; CURRENT APPLICATION NUMBER: US/09/167,681A
; CURRENT FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 8447
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (4361)...(4507)
; NAME/KEY: CDS
; LOCATION: (4612)...(4737)
; NAME/KEY: CDS
; LOCATION: (4827)...(4925)
; NAME/KEY: CDS
; LOCATION: (6322)...(6447)
; NAME/KEY: CDS
; LOCATION: (6543)...(6638)
; NAME/KEY: CDS
; LOCATION: (7137)...(7316)
; NAME/KEY: CDS
; LOCATION: (7439)...(7553)
; US-09-167-681-45

Query Match          7.5%; Score 215.2; DB 4; Length 8453;
Best Local Similarity 85.1%; Pred. No. 4.4e-43;
Matches 252; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 2565 AGCGCGGCTCAGTGGCTCACACCTGTAATCCAGCACTTTGGGAGGCCAAGTGGCGG 2624
Db 2000 AGGCTGAGCGTGGTGGCTCAGCCTGTAATCCAGTACTTTGGGAGGTGCGATGAGCAGG 1941

QY 2625 ATCACTGAGGTGAGGAGTTTGGCA-CAAGCTGCCAACAGCTGAAACCCCATCTCCAC 2683

; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3
Query Match          7.5%; Score 215.8; DB 4; Length 80595;
Best Local Similarity 85.4%; Pred. No. 7.1e-43;
Matches 252; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 2561 CCCAGCGCGGCTCAGTGGCTCACACCTGTAATCCAGCACTTTGGGAGGCCAAGTGG 2620
Db 14598 CTCTGGCCAGGCATGGTGGCTTACGCCCTGTAATCCAGCAGTTTAGGAGGTCGAGTGG 14539

QY 2621 GCGGATCACTGAGTGCAGGAGTTTGCACACGCT-CCCAACAGCTGAAACCCCATCT 2679
Db 14538 GCGGATCACTGAGTGCAGGAGTTTGCACACGCTGCGCCACATGTTGTAACCCCATCT 14479

QY 2680 CCACATAAAATACAAAAATTAGTTGGCATGGTGTGAGCACCTGTATATCCAGCTACTC 2739
Db 14478 CTACAAAAACACAAAAATTAGTTGGCATGGTGTGAGCACCTGTATATCCAGCTACTC 14419

QY 2740 TGGAGGCTGAGATGAGGATCATTGAAACCGGAGGTGGAGTTGCAGTGAGTAAGA 2799
Db 14418 AGGAGGCTGAGCATGAGAACTGCTTGAACCGGAGGCGGAGGTTGCAGTGAGCTGTA 14359

QY 2800 TCACATCACTGCACCTCCAGCCTGGGTAACAGAGTGAGACTGTCTCAAAAAA 2854
Db 14358 TCAGTCACTGCATCCAGCCTAGGCAACAGTGAGACTGCATCTCAAAAAA 14304

RESULT 13
US-09-729-995-3
; Sequence 3, Application US/09729995
; Patent No. 6426206
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00904
; CURRENT APPLICATION NUMBER: US/09/729,995
; CURRENT FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 29629
; TYPE: DNA
; ORGANISM: Human
; US-09-729-995-3

Query Match          7.5%; Score 215.2; DB 4; Length 29629;
Best Local Similarity 83.9%; Pred. No. 6.9e-43;
Matches 255; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 2557 TGTGCCCGCAGCGCGGCTCAGTGGCTCACACCTGTAATCCAGCACTTTGGGAGGCCAAG 2616
Db 11621 TCTGCTCCAGCTGGGCGCGTGTGCTCACGCTGTAATCCAGCACTTTGGGAGGCCGAG 11680

QY 2617 GTGGCGGATCACCTGAGGTGAGGAGTTTGGCAAGCCTG-CCACAAAGCTGAAACCCC 2675
Db 11681 GTGGGTGATCACCTGAGGTGAGGAGTTCAAGACAGCCTGACCAACATGCGAAACCCCT 11740

QY 2676 ATCTCCACTAAATACAAAAATTAGTTGGCATGGTGTGAGCACCTGTAATCCAGCT 2735
Db 11741 GTCTCTACTAAAAATACAAAAAGTAGCGCGCGTGTGGCAGACACCTGTAATCCCACT 11800

QY 2736 ACTCTGGAGGCTGAGATGAGGAGTCACTTGAACCCGGGAGGTGGAGTTGAGTGAGCT 2795
Db 11801 ACTCGGAGGCTGAGGAGGAGGAGAACTCTTGNACCTGGGAGGTGGAGGCTGAGTGAGCC 11860

QY 2796 AAGATCACATCACTGCACCTCCAGCCTGGGTACAGAGTGAGAGTGTCTCAAAAAA 2855
Db 11861 GAGATCACGCCATTGCACTCCAGCCTGGGAGGAGAGAGTGTCTCTCAAAAAA 11920

QY 2856 AAAA 2859
Db 11921 AAAA 11924

RESULT 14
US-09-799-345-1/c
; Sequence 1, Application US/09799345
; Patent No. 6323016
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
```

```
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001156
; CURRENT APPLICATION NUMBER: US/09/799,345
; CURRENT FILING DATE: 2001-03-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Human
US-09-799-345-1

Query Match 7.5%; Score 213.8; DB 4; Length 2532;
Best Local Similarity 80.6%; Pred. No. 6.3e-43;
Matches 262; Conservative 0; Mismatches 62; Indels 1; Gaps 1;

QY 2536 GTGAAGGTGACATCAGGATTTATGTCCTCCAGCGCGGCTCAGTGGCTCACACCTGTAAATC 2595
DB 1544 GTGAATAGAGATAAGATGACTTCTTTAGGCCAGGCAGCGTGGCTCAGCGCTATAATC 1485
QY 2596 CCAGCACTTTGGAGGCCAAGGTGGCGGATCAGTGTGAGGTCAGGAGTTTCCGACAGCC 2655
DB 1484 CCAGCACTCTGGAAGGCCAGGTAGTGGATCAGTGTGAGGTCAGGAGTTTGGAGACCAACC 1425
QY 2656 T-GCCAAACAGCTGAACCCCATCTCCACTAAAAATACAAAAATAGTTGGGATGGTG 2714
DB 1424 TGGCCAAAGTGTGAACCCCGCTCTACTAAAAATACAAAAATAGCTGGGATGGTG 1365
QY 2715 TGAGCACCTGTAATCCAGCTACTCTGGAGGCTGAGATGAGGATGATCACTTGAACCCGGG 2774
DB 1364 TGCACGCTATAATCCAGCCACTCAGGAGGCTGAGGAGGAGAAATGCTTGAACCCAGG 1305
QY 2775 AGTGGAGGTTCAGTGTGAGTAAAGATCACATCACTGCACATCCAGCCTGGGTAACAGAGTG 2834
DB 1304 AGTGGAGGTTCAGTGTGAGTAAAGATCAAGCCAGTGCACATCCAGCCTGGGTAACAGAGCA 1245
QY 2835 AGACTGTCTCAAAAAAATAAAAAA 2859
DB 1244 AGAGAGACTCTTATTAAAAAANA 1220

RESULT 15
US-08-068-945A-1
; Sequence 1, Application US/08068945A
; Patent No. 5616483
; GENERAL INFORMATION:
; APPLICANT: Bjursell, Gunnar
; APPLICANT: Carlsson, Peter
; APPLICANT: Enerback, Sven
; APPLICANT: Hansson, Lennart
; APPLICANT: Lidberg, Ulf
; APPLICANT: Nilsson, Jeanette
; APPLICANT: Tornell, Jan
; TITLE OF INVENTION: New DNA Sequences
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/068,945A
; FILING DATE: 27-MAY-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9201809-2
; FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9201826-6
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9202088-2
; FILING DATE: 03-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300902-5
; FILING DATE: 19-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)819-8783
; TELEFAX: (212)354-8113
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11531 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Mammary gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1653..1727, 4071..4221, 4307..4429, 4707
; LOCATION: ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
; LOCATION: ..8521, 8719..8922, 10124..10321, 10650..11391)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: join(1722..1727, 4071..4221, 4307..4429, 4707
; LOCATION: ..4904, 6193..6323, 6501..6608, 6751..6868, 8335
; LOCATION: ..8521, 8719..8922, 10124..10321, 10650..11391)
; OTHER INFORMATION: /EC_number= 3.1.1.1
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..1640
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 1611..1617
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1641..1727
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4071..4221
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4307..4429
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4707..4904
; FEATURE:
; NAME/KEY: exon
; LOCATION: 6193..6323
; FEATURE:
; NAME/KEY: exon
; LOCATION: 6501..6608
; FEATURE:
; NAME/KEY: exon
; LOCATION: 6751..6868
; FEATURE:
; NAME/KEY: exon
; LOCATION: 8335..8521
; FEATURE:
; NAME/KEY: exon
; LOCATION: 8719..8922
; FEATURE:
```

```

; NAME/KEY: exon
; LOCATION: 10124...10321
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10650...11490
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 11491...11531
; US-08-068-945A-1

Query Match      7.4%; Score 212.6; DB 1; Length 11531;
Best Local Similarity 84.7%; Pred. No. 2.1e-42; Indels 1; Gaps 1;
Matches 250; Conservative 0; Mismatches 44;

QY 2566 GCGCGGCTCAGTGGCTCACACCTGTAAATCCAGCACCTTTGGAGGCCAAGGTGGCGGGA 2625
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1971 GCGCGGCGCAGTGGCTCACACCTGTCAATCCACCAGTTTGGAGGCCGAGGTGGGTGA 2030
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2626 TCACCTGAGGTGAGGAGTTTGGCGACAAGCCT-GCCACACAGCTGAACCCCATCTCCACT 2684
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2031 TCACTTGAGGTGAGGAGTTGGAGACCAGCCTGGGCCAACATGGTGAACCCCGTCTCTACT 2090
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2685 AAAAATACAAAAATTAGTTGGGCATGGTGTGACACCTGTAAATCCAGCTACTCTGGAG 2744
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2091 AAAAATACAAATATTAATCGGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2150
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2745 GCTGAGATAGGAGGATCACTTTGAACCGGGAGGTGGAGGTTCAGTGCAGTAAAGATCACA 2804
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2151 GCTGAGGCAGGAACCGCTTGAAGCTGGGAGGTGGAGATTGGCGGTGAGCTGAGATCAGC 2210
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2805 TCACCTGCACCTCAGCCTGGTAAACAGAGTGAAGTGTCTCAAAAAAAGAAAAA 2859
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2211 CTACTGCACCTCAGCCTGGGTGACAGGCGGAGACTCTCTCAAAAAATAGAAAA 2265
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: July 17, 2003, 21:01:41
Job time : 138 secs

QY	2461	TTGCACCGAGACCTTAAC	TCGCGCTCAGAGTG	TAGACCCGAGATGGTGCAGATGCC	TCGAGT	2520
Db	2461	TTGCACCGAGACCTTAAC	TCGCGCTCAGAGTG	TAGACCCGAGATGGTGCAGATGCC	TCGAGT	2520
QY	2521	GCCATTAAANTGTGGTGA	AGGTCAGATCAGGAT	TATGTGCCCGAGCCGGGCTCAGTGG		2580
Db	2521	GCCATTAAANTGTGGTGA	AGGTCAGATCAGGAT	TATGTGCCCGAGCCGGGCTCAGTGG		2580
QY	2581	CTCACACCTGTAAATCCC	AGCACTTTGGGAGGCCAAG	TGGCGGATCACCTGAGGTCAGG		2640
Db	2581	CTCACACCTGTAAATCCC	AGCACTTTGGGAGGCCAAG	TGGCGGATCACCTGAGGTCAGG		2640
QY	2641	AGTTTGCACAAAGCCTGC	CAACAAAGCTGAAACCCCA	TCTCCACTTAAANAATACAAA	AATTA	2700
Db	2641	AGTTTGCACAAAGCCTGC	CAACAAAGCTGAAACCCCA	TCTCCACTTAAANAATACAAA	AATTA	2700
QY	2701	GTTGGGCATGTGTGAGCA	CACTGTAAATCCCAAGTCT	TCTGGAGGCTGAGATAGGAGGAT		2760
Db	2701	GTTGGGCATGTGTGAGCA	CACTGTAAATCCCAAGTCT	TCTGGAGGCTGAGATAGGAGGAT		2760
QY	2761	CACTTGAACCCGGGAGGT	TGGAGTTGCAGTGAAGTAC	ATACATCACATCACTGCAC	TCCAGCC	2820
Db	2761	CACTTGAACCCGGGAGGT	TGGAGTTGCAGTGAAGTAC	ATACATCACATCACTGCAC	TCCAGCC	2820
QY	2821	TGGGTAACACAGATGAGAC	TCTCTCAAAAAAAAAAAAAA			2859
Db	2821	TGGGTAACACAGATGAGAC	TCTCTCAAAAAAAAAAAAAA			2859

RESULT 2

US-09-923-684-1

; Sequence 1, Application US/0992368A

Patent No. US20020081613A1

; GENERAL INFORMATION:

APPLICANT: Narayanan, Ramaswamy

; TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER

FILE REFERENCE: 6818-24

; CURRENT APPLICATION NUMBER: US/09/9

2001-09-17

; NUMBER OF SEQ ID NOS: 16

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1

```

; SEV ID NO I
; LENGTH: 3920

LENGTH: 3920
TYPE: DNA

TYPE: DNA
ORGANISM: Homo sapiens

US-09-923-684-1

Query Match 58.1%; Score 1661.6; DB 10; Length 3920;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1664; Conservative 0; Mismatches 4; Total 4

Qy	1	ACTCACTATAGGGCTTCAGCGCGCCCGCGCAGGTGGGCTTCCCGGGCCCTGGAGCAGC	60
Db	1	ACTCACTATAGGGCTTCAGCGCGCCCGCGCAGGTGGGCTTCCCGGGCCCTGGAGCAGC	60
Qy	61	GCCGGGCTTAATATGCCCGGAGCGGCGCGATGAAGAGAAGTCCAGAATCGCGCCA	120
Db	61	GCCGGGCTTAATATGCCCGGAGCGGCGCGATGAAGAGAAGTCCAGAATCGCGCCA	120
Qy	121	AGACCAGGAGGAGAAGAAATGGCGAGTTTACGAGTTGGCCAAGTGTCTCCCGCTGC	180
Db	121	AGACCAGGAGGAGAAGAAATGGCGAGTTTACGAGTTTGGCAAGTGTCTCCCGCTGC	180
Qy	181	CGTGGCCCATCACTTCGACGCTGGACAAAGCGTCCATCATCGGCTCACACAGAGTACC	240
Db	181	CGTGGCCCATCACTTCGACGCTGGACAAAGCGTCCATCATCGGCTCACACAGAGTACC	240
Qy	241	TGAAGATCGCGCGCTTTCGCCGAAGTTTATAGGAGCGCGTGGGGACAGCCGAGCGCG	300
Db	241	TGAAGATCGCGCGCTTTCGCCGAAGTTTATAGGAGCGCGTGGGGACAGCCGAGCGCG	300
Qy	301	CCGGGCCCTTGAGCGCGTTCGCCAAGGAGCTGGGATTCGCACTTTCGCTCGACACTTTGGATG	360

Db 1381 TTCTGTACAGCCATCCTACAGCGTGCCTTCTCTACCATACGGACACTTCCCTCTGG 1440

Qy 1441 ACTCTCACTTCTCAGCAGCAAAAGCAATGTTGGCGCAAGTTCTGGGACGCCCAAG 1500

Db 1441 ACTCTCACTTCTCAGCAGCAAAAGCAATGTTGGCGCAAGTTCTGGGACGCCCAAG 1500

Qy 1501 GATCCCTTGTGAGGTGGCAGCGCTTTTCTCTAGCACAATGCCAGCAGCGGTGAATGCC 1560

Db 1501 GATCCCTTGTGAGGTGGCAGCGCTTTTCTCTAGCACAATGCCAGCAGCGGTGAATGCC 1560

Qy 1561 AGTGGCATTATGCCAACCCCTAGTGTCTAGCAGCTGTCTCCAGCTTAAATAATCTCCAG 1620

Db 1561 AGTGGCATTATGCCAACCCCTAGTGTCTAGCAGCTGTCTCCAGCTTAAATAATCTCCAG 1620

Qy 1621 AGCCACCGGCAACTCTAGGCAAGCGCTGTGCGCAAGCTACGAAG 1668

Db 1621 AGCCACCGGCAACTCTAGGCAAGCGCTGTGCGCAAGCTACGAAG 1668

RESULT 3

US-09-923-684-10

Sequence 10, Application US/09923684

Patent No. US20020081613A1

GENERAL INFORMATION:

APPLICANT: Narayanan, Ramaswamy

TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER

FILE REFERENCE: 6818-24

CURRENT APPLICATION NUMBER: US/09/923,684

CURRENT FILING DATE: 2001-09-17

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn version 3.1

SEQ ID NO 10

LENGTH: 472

TYPE: DNA

ORGANISM: Homo sapiens

US-09-923-684-10

Query Match 16.3%; Score 465.6; DB 10; Length 472;

Best Local Similarity 99.2%; Pred. No. 6.1e-120;

Matches 468; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1158 TGGAGGACCGCCTGTCTACCTCACAAGAACTAGGAATTTAGTGAACCCCAAAATACC 1217

Db 1 TGGAGGACCGCCTGTCTACCTCACAAGAACTAGGAATTTAGTGAACCCCAAAATACC 60

Qy 1218 AAGATGAGCAAAAGCTGAGCAAAACCTTACCCCCACAGCAATACAGCTCATTCCAA 1277

Db 61 AAGATGAGCAAAAGCTGAGCAAAACCTTACCCCCACAGCAATACAGCTCATTCCAA 120

Qy 1278 ATGGACAACTGGAATCGGCCAGCTCGGAACTGGAGAGCCAGTCCCTCGAAGCGCT 1337

Db 121 ATGGACAACTGGAATCGGCCAGCTCGGAACTGGAGAGCCAGTCCCTCGAAGCGCT 180

Qy 1338 GCTGCTCCTCAGAACTGAGCCCACTCAGAAAGCAGTGACCTTCTGTACAGGCCATCC 1397

Db 181 GCTGCTCCTCAGAACTGAGCCCACTCAGAAAGCAGTGACCTTCTGTACAGGCCATCC 240

Qy 1398 TACAGCTGCGCTTCTCTACCAATATGAGCACTTCCCTGAGCACTCTCACTTCTTCAGC 1457

Db 241 TACAGCTGCGCTTCTCTACCAATATGAGCACTTCCCTGAGCACTCTCACTTCTTCAGC 300

Qy 1458 AGCAAAAGCAATGTTGGCGGCAAGTTGGGAGCCCAAGGATCCCTCTGTGAGGTG 1517

Db 301 AGCAAAAGCAATGTTGGCGGCAAGTTGGGAGCCCAAGGATCCCTCTGTGAGGTG 360

Qy 1518 GCAGCTTTTCTTCTGAGCAATGCCAGCCAGCGGTGAATGCCAGTGGCAATTATGCCAAC 1577

Db 361 GCAGCTTTTCTTCTGAGCAATGCCAGCCAGCGGTGAATGCCAGTGGCAATTATGCCAAC 420

Qy 1578 CCCCTAGTGCCTAGCAGCTCTCTCAGCTTAAATAATCTCCAGAGCCACCGG 1629

Db 421 CCCCTAGTGCCTAGCAGCTCTCTCAGCTTAAATAATCTCCAGAGCCACCGG 472

RESULT 4

US-09-902-214-6

Sequence 6, Application US/09902214

Publication No. US20030104521A1

GENERAL INFORMATION:

APPLICANT: Whitaker, Paul Andrew

TITLE OF INVENTION: Disease-Associated Gene

FILE REFERENCE: 4-31503A/H031

CURRENT APPLICATION NUMBER: US/09/902,214

CURRENT FILING DATE: 2001-07-10

NUMBER OF SEQ ID NOS: 84

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 6

LENGTH: 50000

TYPE: DNA

ORGANISM: Homo sapiens

US-09-902-214-6

Query Match 8.1%; Score 230.6; DB 12; Length 50000;

Best Local Similarity 80.2%; Pred. No. 1.9e-53;

Matches 283; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

Qy 2508 AGATGCTCAGTGCCTTAAATGTTGGTGAAGTGACATCAGGATTATGTGCCCCAGG 2567

Db 40656 AAATGCCAGAAAGAACTGCTTAAAGGTGTCACAGTGCCTCAAGAAGTATGTGACTGG 40715

Qy 2568 CCGGGCTCAGTGCCTCAGCCTTAATCCAGCATTTCGGAGGCCAAGTGGGCGGATC 2627

Db 40716 CCGGGCTCAGTGCCTCAGCCTTAATCCAGCATTTCGGAGGCCAAGTGGGCGGATC 40775

Qy 2628 ACCTGAGTCAAGATTGGCACAAGCT-GCCACAAGCTGAAACCCCATCTCCACTAA 2686

Db 40776 ACCTGAGTCAAGATTGGCACAAGCT-GCCACAAGCTGAAACCCCATCTCCACTAA 40835

Qy 2687 AATACAAAAATTAGTTGGGCATGGTGTGAGCAGCTGTATATCCAGCTACTCTGAGGC 2746

Db 40836 AGATACAAAAATTAGTTGGGCATGGTGTGAGCAGCTGTATATCCAGCTACTCTGAGGC 40895

Qy 2747 TGAGATAGGAGTCACTTGAACCCGGAGGTGAGCTTCAAGTGAAGTCAATC 2806

Db 40896 TGAGATAGGAGTCACTTGAACCCGGAGGTGAGCTTCAAGTGAAGTCAATC 40955

Qy 2807 ACTCACTCCAGCTGGTGAACAGAGTGAGCTGTCTCAAAAAA 2859

Db 40956 ACTCACTCCAGCTGGTGAACAGAGTGAGCTGTCTCAAAAAA 41008

RESULT 5

US-10-027-632-139070/c

Sequence 139070, Application US/10027632

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 139071
LENGTH: 752
TYPE: DNA
ORGANISM: Human
US-10-027-632-139071

Query Match 7.9%; Score 226.4; DB 15; Length 752;
Best Local Similarity 87.5%; Pred. No. 5.7e-53;
Matches 259; Conservative 0; Mismatches 36; Indels 1; Gaps 1;

QY 2564 CAGCCCGGCTCAGTGGCTCACACCTGTATCCAGCAGCTTTGGGAGGCCAGGTGGCG 2623
DB 102 CTGGCCGGCGCGGTGCTCATGCTGTATCCAGCAGCTTTGGGAGGCCAGGTGGCG 161
QY 2624 GATCAGCTGAGTCAGGAGTTTGGCGAAGCCTT-GCCAAAGCTGAAACCCCATCTCCA 2682
DB 162 GATCAGCTGAGTCAGGAGTTTGGCGAAGCCTTGGCCAGCAGCTGTAACCCCGTCTCTA 221
QY 2683 CTAATAATACAAAAATTAGTTGGCAGTGGTGGAGCAGCTGTAAATCCAGCTACTCTGG 2742
DB 222 CTAATAATACAAAAATTAGTTGGCAGTGGTGGAGCAGCTGTAAATCCAGCTACTCTGG 281
QY 2743 AGGCTGAGTAGGAGGATCACTTGAACCGCGGAGGTGGAGGTTCAGTGAAGATCA 2802
DB 282 AGGCTGAGTAGGAGGATCACTTGAACCGCGGAGGTGGAGGTTCAGTGAAGATCA 341
QY 2803 CATCACTGCACTCCAGCCTGGGTAAACAGAGTGAGACTGTCTCAAAAAAAGAAAAA 2858
DB 342 CACCACTGCACTCCAGCCTGGGTAAACAGAGTGAGACTGTCTCAAAAAAAGAAAAA 397

RESULT 10
US-10-027-632-264542
Sequence 264542, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23

524 GTTCAGTGGCCAGATCACACCTGCACTCCAGCCTGGGTGACAGAGTGAGACTCG 583
QY 2843 TCAAAAAAAGAAAAA 2859
DB 584 TCTCAAAAAATAAATA 600

RESULT 8
US-10-027-632-287603
Sequence 287603, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287603
LENGTH: 489
TYPE: DNA
ORGANISM: Human
US-10-027-632-287603

Query Match 7.9%; Score 226.6; DB 15; Length 489;
Best Local Similarity 82.6%; Pred. No. 4.2e-53;
Matches 270; Conservative 1; Mismatches 55; Indels 1; Gaps 1;

QY 2534 GGTGAGGTGACATCAGGATTTATGCCCCAGCGCGGTGCTCAGTGGCTCACACCTGTAA 2593
DB 148 GGGGAAAGTCCCATCAGAAACATCTAACACAGCGCGGCGGCTCATGCTGTAA 207
QY 2594 TCCAGCAGCTTTGGGAGGCCAAGTGGGGGATCACTGAGTCAGGAGTTTGGCACAAG 2653
DB 208 TCCAGCAGCTTTGGGAGGCCAAGTGGGGGATCACTGAGTCAGGAGTTTGGCACAAG 267
QY 2654 CCT-GCCAAAGCTGAAACCCCATCTCCACTAAATAACAAAAATTAGTTGGCATGT 2712
DB 268 CTGGCCCAACATGTTGAAACCCCGTCTACTATAAATTACAAAAATTAGTTGGCTGT 327
QY 2713 GGTGAGCAGCTGTAATCCAGCTACTCTGGAGGTGAGATGAGGAGTCACTTTGAACCCG 2772
DB 328 GCGGCGCCCTGTAATCCAGCTACTCAGGAGGTGAGGAGGAGTAATGCTTTGAACCCG 387
QY 2773 GGAGTGAGGTTGCAGTGAGCTAAGATCATCATCTGCACTCCAGCCTGGTAACAGAG 2832
DB 388 GGAGCGGAGTTGTCAGTGAGCGGAGCACGACCTGCACTCCAGCCTGGGCAACAG 447
QY 2833 TGAGACTGTCTCAAAAAAAGAAAAA 2859
DB 448 CAAGACTGTCTCAAAAAAAGAAAAA 474

RESULT 9
US-10-027-632-139071
Sequence 139071, Application US/10027632
GENERAL INFORMATION:

	Query Match	7.9%	Score 225.2;	DB 11;	Length 10612;
	Best Local Similarity	85.4%;	Pred. No. 3.4e-52;		
	Matches 251;	Conservative	0;	Mismatches 43;	Indels 0;
	Gaps	0;			
QY	2566	GGCCGGGCTCAGTGGCTCAGCACTGTAATCCAGCACATTTGGGAGGCCAAGGTGGCGGA	2625		
DB	2259	GGCTGGCGCAGTGGCTCAGCCCTGTATATCCAGCACATTTGGGAGGCCAAGGCGGTTGGA	2200		
QY	2626	TCACCTGAGGTCAGGAGTGTTCGACAAGCCTTGCCACAACAGCTGAAACCCCACTCTCCACTA	2685		
DB	2199	TGTGCTTGAGGTCAGAAGTTGCGATCAGCCTGGCCAAAAGGTGAAACCCCTTTTCTACTA	2140		
QY	2686	AAATACAAAAAATTAGTTGGGCATGTTGGTGAGCACTGTAATCCAGCTACTCTGGAGG	2745		
DB	2139	AAATACAAAAAATTAGCTGGCCCTGGTGGTGGCTGTGCTGTGCTGTAAATTCACGCTACTCAGAGG	2080		
QY	2746	CTGAGATAGGAGGATCACTTGAACCCGGAGGTGGAGGTTGCGAGTGAGCTAAGATCATCAT	2805		

```

RESULT 13
US-09-764-877-2651
; Sequence 2651, Application US/09764877
; Patent No. US2002014140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2651
; LENGTH: 11337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2651

```

Query Match 7.8%; Score 224.4; DB 11; Length 11337;
Best Local Similarity 87.4%; Pred. No. 5.8e-52;

Mon Jul 21 15:46:56 2003

us-09-923-684-2.rnpb

Matches 257; Conservative 0; Mismatches 36; Indels 1; Gaps 1;
QY 2567 GCCGGCTCAGTGGCTCACACCTGTAATCCAGCAGCTTTGGGAGGCAAGGTGGGGGAT 2626
Db 2315 GCCAGGCGGTGCTCATGCTGTAATCCAGGACTTTGGGGGCGGAGGTGGGTGGAT 2374
QY 2627 CACCTGAGGTGAGGAGTGGGACAGCCCT-GCCAAACAGCTGAACCCCTCTCCACTA 2685
Db 2375 CACCTGAGGTGAGGAGTTCGAGACAGCCCTGGCCCAACATGTTGAACCCCTCTCTACTA 2434
QY 2686 AAAATACAAAATTAGTTGGGATGGTGGTGGACACCTGTAATCCAGCTACTCTGGAGG 2745
Db 2435 AAAATGCAAAAATTAGCGGCGGTGGTGGTGCACACCTGTAGTCCAGCTACTCAGGAGG 2494
QY 2746 CTGAGATAGGAGTACATTTGAACCCGGGAGGTGGAGTTGCCAGTGAAGTAAATCAT 2805
Db 2495 CTGAGGAGGAGATCATTTGAACCCGGGAAAGTGGAGTTGCCAGTGAAGGATTCAC 2554
QY 2806 CACTGCACCTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAAAAAA 2859
Db 2555 CACTGCACCTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAAAAAA 2608

RESULT 14
US-10-017-128-1/c
; Sequence 1, Application US/10017128
; Publication No. US2003012436A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT FOR VASCULAR DISEASE
; FILE REFERENCE: MMI-001
; CURRENT APPLICATION NUMBER: US/10/017,128
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/306,941
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/315,572
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/327,488
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-128-1

Query Match 7.8%; Score 224.2; DB 15; Length 146547;
Best Local Similarity 84.3%; Pred. No. 1.7e-51;
Matches 264; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 2548 TCAGGATTATGTGCCCCAGCGGGCTCAGTGGCTCACACCTGTAATCCAGCAGCTTTGG 2607
Db 75123 TCAAAATTAACGTGACTTGGCTGGCGCAGTGGCTCACACCTGTAATCCAGCAGCTTTGG 75064
QY 2608 GAGGCCAAGGTGGCGGATCACCTGAGGTGAGGAGTTGGGACAAAGCTTG-CCAAACAGC 2666
Db 75063 GAGGCCAAGGTGGCGGATCACCTGAGGTGAGGAGTTGGGACAAAGCTTG-CCAAACAGC 75004
QY 2667 TGAACCCCATCTCCACTAAATAACAAATAGTTGGGATGGTGGTGGAGCACCTGTA 2726
Db 75003 TGAACCCCATCTCTACTAAATAACAAATAGTTGGGATGGTGGTGGAGCACCTGTA 74944
QY 2727 ATCCAGCTACTCTGAGGCTGAGATAGGAGGATCACTTGAACCCGGGAGGTGGAGTTG 2786
Db 74943 ATCCAGCTACTCTGAGGCTGAGATAGGAGGATCACTTGAACCCGGGAGGTGGAGTTG 74884
QY 2787 CAGTGAGTAAATCACTCACTCCAGCTGGTAAACAGAGTGAGACTGTCTCAA 2846
Db 74883 CAGTGAGTAAATCACTCACTCCAGCTGGTAAACAGAGTGAGACTGTCTCAA 2846
QY 2847 AAAAAAAAAAAAAA 2859
Db 74884 CAGTGAGTAAATCACTCACTCCAGCTGGTAAACAGAGTGAGACTGTCTCAA 74824

Db 74823 AAAAAAAAAAAAAA 74811
RESULT 15
US-09-764-891-6056/c
; Sequence 6056, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6056
; LENGTH: 1743
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6056
Query Match 7.8%; Score 223.8; DB 12; Length 1743;
Best Local Similarity 87.1%; Pred. No. 4.2e-52;
Matches 257; Conservative 0; Mismatches 37; Indels 1; Gaps 1;
QY 2566 GGCGGGCTCAGTGGCTCACACCTGTAATCCAGCAGCTTTGGGAGGCAAGGTGGCGGA 2625
Db 409 GGCTGGGTGCTGGCTCACACCTATAATCCAGCAGCTTTGGGAGGCAAGGTGGCGGA 350
QY 2626 TCACCTGAGGTGAGGAGTTTGGGACAAAGCCT-GCCAAACAGCTGAACCCCTCTCCACT 2684
Db 349 TCACCTGAGGTGAGGAGTTTCAAGACAGCCTGGCCAAACATGGTGAACCCCTCTCTACT 290
QY 2685 AAAAAATACAAAATTAGTTGGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2744
Db 289 AAAAAATACAAAATTAGTCCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 230
QY 2745 GCTGAGATAGGAGTCACTTGAACCCGGGAGGTGGAGTTGTCAGTGGAGTAAAGATCACA 2804
Db 229 GCTGAGGAGGAGGAGTCACTTGAACCCCTGGGAGGTGGAGTTGTCAGTGGAGTCAAGATCGC 170
QY 2805 TCACCTGCACTCCAGCTGGGTAAACAGAGTGAGACTGTCTCAAAAAA 2859
Db 169 CCACTGCACTCCAGCTGGGTGACAGAGTGAGACTGTCTCTTAAAAA 115

Search completed: July 17, 2003, 21:11:56
Job time : 607 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 17:52:52 ; Search time 1857 Seconds
(without alignments)
376.127 Million cell updates/sec

Title: US-09-923-684-12
Perfect score: 24
Sequence: 1 gagagcaagaagacacagcaagcc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 521186

Minimum DB seq length: 0
Maximum DB seq length: 25

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl: *
- 1: gb_ba:*
 - 2: gb_htg:*
 - 3: gb_in:*
 - 4: gb_om:*
 - 5: gb_ov:*
 - 6: gb_pat:*
 - 7: gb_ph:*
 - 8: gb_pl:*
 - 9: gb_pr:*
 - 10: gb_ro:*
 - 11: gb_sts:*
 - 12: gb_sy:*
 - 13: gb_un:*
 - 14: gb_vl:*
 - 15: em_ba:*
 - 16: em_fun:*
 - 17: em_hum:*
 - 18: em_in:*
 - 19: em_mu:*
 - 20: em_om:*
 - 21: em_or:*
 - 22: em_ov:*
 - 23: em_pat:*
 - 24: em_ph:*
 - 25: em_pl:*
 - 26: em_ro:*
 - 27: em_sts:*
 - 28: em_un:*
 - 29: em_vl:*
 - 30: em_htg_hum:*
 - 31: em_htg_inv:*
 - 32: em_htg_other:*
 - 33: em_htg_mus:*
 - 34: em_htg_pln:*
 - 35: em_htg_rod:*
 - 36: em_htg_rnd:*
 - 37: em_htg_vrt:*
 - 38: em_sy:*
 - 39: em_htgo_hum:*
 - 40: em_htgo_mus:*
 - 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	16	66.7	25	6	AX117664	AX117664 Sequence
2	15.2	63.3	21	6	AX452262	AX452262 Sequence
C 3	15.2	63.3	20	6	AX288411	AX288411 Sequence
C 4	15	62.5	24	6	AX293044	AX293044 Sequence
5	14.4	60.0	25	6	AX429796	AX429796 Sequence
6	14.2	59.2	20	6	AX191307	AX191307 Sequence
C 7	13.8	57.5	17	6	AX475389	AX475389 Sequence
C 8	13.8	57.5	25	6	AX476366	AX476366 Sequence
C 9	13.8	57.5	25	6	AX476367	AX476367 Sequence
C 10	13.8	57.5	25	6	AX476368	AX476368 Sequence
C 11	13.8	57.5	25	6	AX476369	AX476369 Sequence
C 12	13.8	57.5	25	6	AX476370	AX476370 Sequence
C 13	13.8	57.5	25	6	AX476371	AX476371 Sequence
C 14	13.8	57.5	25	6	AX476372	AX476372 Sequence
C 15	13.8	57.5	25	6	AX476373	AX476373 Sequence
C 16	13.8	57.5	25	6	AX476374	AX476374 Sequence
C 17	13.6	56.7	22	6	AR207807	AR207807 Sequence
18	13.4	55.8	20	4	BOVDIK08	D44509 Bovine DNA,
19	13.4	55.8	22	6	AX021010	AX021010 Sequence
20	13.4	55.8	24	6	AR066891	AR066891 Sequence
C 21	13.2	55.0	22	6	AR066609	AR066609 Sequence
C 22	13.2	55.0	23	6	I34867	I34867 Sequence 10
C 23	13	54.2	21	6	A26766	A26766 Oligonucleo
24	13	54.2	21	6	AR004454	AR004454 Sequence
C 25	12.8	53.3	17	6	AX475388	AX475388 Sequence
C 26	12.8	53.3	17	6	AX475390	AX475390 Sequence
C 27	12.8	53.3	20	12	AB068351	AB068351 Synthetic
C 28	12.8	53.3	21	6	AR066947	AR066947 Sequence
C 29	12.8	53.3	23	6	AX487008	AX487008 Sequence
C 30	12.8	53.3	24	6	AX164356	AX164356 Sequence
C 31	12.8	53.3	25	6	AX179195	AX179195 Sequence
C 32	12.8	53.3	25	6	AX476365	AX476365 Sequence
C 33	12.8	53.3	25	6	AX476375	AX476375 Sequence
34	12.6	52.5	22	6	AX033431	AX033431 Sequence
35	12.4	51.7	25	6	AR137526	AR137526 Sequence
36	12.4	51.7	25	6	AR139032	AR139032 Sequence
37	12.4	51.7	25	6	AX300826	AX300826 Sequence
38	12.2	50.8	20	6	A57378	A57378 Sequence 15
39	12.2	50.8	20	6	A61160	A61160 Sequence 1
C 40	12.2	50.8	20	6	A83993	A83993 Sequence 14
41	12.2	50.8	20	6	AR020539	AR020539 Sequence
C 42	12.2	50.8	20	6	AR089400	AR089400 Sequence
C 43	12.2	50.8	20	6	AR093600	AR093600 Sequence
C 44	12.2	50.8	20	6	AX030695	AX030695 Sequence
45	12.2	50.8	20	6	AX047253	AX047253 Sequence

ALIGNMENTS

RESULT 1
AX117664
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX117664
Sequence 2787 from Patent WO0129262.
AX117664
AX117664.1 GI:14034615
synthetic construct.
artificial construct.
artificial sequences.
1 (bases 1 to 25)
Picoult-Newburg, L. and Pohl, M.
Genotyping reagents, kits and methods of use thereof
Patent: WO 0129262-A 2787 26-APR-2001;
Orchid Biosciences, Inc. (US)

Linear PAT 11-MAY-2001

```
FEATURES
source
1. .25
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"
10 a 8 c 6 g 1 t

BASE COUNT
10 a 8 c 6 g 1 t
ORIGIN

Query Match
66.7%; Score 16; DB 6; Length 25;
Best Local Similarity 79.2%; Pred. No. 5.9e+03;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAGAGCAAGAAAGCAGCAAGCC 24
||||| ||||| || |||||
DB 1 GAGAGCAGCAAGAAAGCAGCAAGCC 24

RESULT 2
AX452262 21 bp DNA linear PAT 06-JUL-2002
LOCUS
Sequence 19 from Patent WO0242442.
DEFINITION
AX452262
ACCESSION
AX452262
VERSION
AX452262.1 GI:21712192
KEYWORDS
synthetic construct.
SOURCE
synthetic construct
artificial sequences.
ORGANISM
1
REFERENCE
Grabowski, R., Braunschweiler, M., Gasch, A. and Berghof, K.
AUTHORS
Novel yeast strain for consumption
TITLE
Patent: WO 0242442-A 19 30-MAY-2002;
JOURNAL
Bioteccon Diagnostics GmbH (DE)
FEATURES
Location/Qualifiers
1. .21
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer zur Differenzierung phylogenetischen
Einheiten, wie Stammen, Unterstammen, Spezies"
7 a 9 c 4 g 1 t

BASE COUNT
7 a 9 c 4 g 1 t
ORIGIN

Query Match
63.3%; Score 15.2; DB 6; Length 21;
Best Local Similarity 85.0%; Pred. No. 1.5e+04;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAAGAAAGCAGCAAGCC 24
||||| ||||| ||||| |||||
DB 2 GCAAGAAAGCAGCAAGCC 21

RESULT 3
AX288411/c 24 bp DNA linear PAT 21-NOV-2001
LOCUS
Sequence 173 from Patent WO0179548.
DEFINITION
AX288411
ACCESSION
AX288411
VERSION
AX288411.1 GI:17050094
KEYWORDS
synthetic construct.
SOURCE
synthetic construct
artificial sequences.
ORGANISM
1
REFERENCE
Barany, F., Zirvi, M., Gerry, N.P., Favis, R. and Kliman, R.
AUTHORS
Method of designing addressable array for detection of nucleic acid
TITLE
sequence differences using ligase detection reaction
JOURNAL
Patent: WO 0179548-A 173 25-OCT-2001;
CORNELL RESEARCH FOUNDATION, INC. (US)
FEATURES
Location/Qualifiers
1. .24
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Hypothetical Probe Sequence"
1 a 9 c 4 g 10 t

BASE COUNT
1 a 9 c 4 g 10 t
ORIGIN

Query Match
62.5%; Score 15; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GAAAGCAGCAAGCC 23
||||| ||||| ||||| |||||
DB 18 GAAAGCAGCAAGCC 4

RESULT 5
AX429796 25 bp DNA linear PAT 21-JUN-2002
LOCUS
Sequence 24 from Patent EP1203826.
DEFINITION
AX429796
ACCESSION
AX429796
VERSION
AX429796.1 GI:21540972
KEYWORDS
synthetic construct.
SOURCE
synthetic construct
artificial sequences.
ORGANISM
1
REFERENCE
Ishizuka, T., Ishiguro, T. and Saitoh, J.
AUTHORS
Oligonucleotide for detection of hiv-1 and detection method
TITLE
Patent: EP 1203826-A 24 08-MAY-2002;
Tosoh Corporation (JP)
JOURNAL
Location/Qualifiers
1. .25
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"
12 a 6 c 6 g 1 t

BASE COUNT
12 a 6 c 6 g 1 t
ORIGIN

Query Match
60.0%; Score 14.4; DB 6; Length 25;
Best Local Similarity 93.8%; Pred. No. 3.6e+04;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 AGAAAGCAGCAAGCC 23
||||| ||||| ||||| |||||
DB 1 AAAAGCAGCAAGCC 16
```


RESULT 6
AX191307
LOCUS AX191307 20 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 4 from Patent WO0149880.
ACCESSION AX191307
VERSION AX191307.1 GI:15209558
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 20)
AUTHORS Korfhage, C. and Oelmüller, U.
TITLE Primers, in particular, for primer-dependent nucleic acid synthesis
JOURNAL processes and nucleic acid amplification methods
Patent: WO 0149880-A 4 12-JUL-2001;
QIAGEN GmbH (DE)
FEATURES
source Location/Qualifiers
1..20
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="n/a"
BASE COUNT 7 a 4 c 8 g 1 t
ORIGIN
Query Match 59.2%; Score 14.2; DB 6; Length 20;
Best Local Similarity 84.2%; Pred. No. 4.5e+04;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 GAGCAAGAGGACAGCAA 21
||||| ||| ||||| |||||
Db 1 GAGCGGAGGACAGCAA 19
||||| ||| ||||| |||||
RESULT 7
AX475389/c
LOCUS AX475389 17 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 610 from Patent WO0224750.
ACCESSION AX475389
VERSION AX475389.1 GI:22214674
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 610 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source Location/Qualifiers
1..17
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 0 a 5 c 5 g 7 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 17;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
||||| ||| ||||| |||||
Db 17 AGAAAGCAGCAGCAGCC 1
||||| ||| ||||| |||||
RESULT 8
AX476366/c
LOCUS AX476366 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1587 from Patent WO0224750.
ACCESSION AX476366
VERSION AX476366.1 GI:22215651
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1587 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source Location/Qualifiers
1..25
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 0 a 10 c 5 g 10 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
||||| ||| ||||| |||||
Db 25 AGAAAGCAGCAGCAGCC 9
||||| ||| ||||| |||||
RESULT 9
AX476367/c
LOCUS AX476367 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1588 from Patent WO0224750.
ACCESSION AX476367
VERSION AX476367.1 GI:22215652
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1588 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source Location/Qualifiers
1..25
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 0 a 9 c 6 g 10 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
||||| ||| ||||| |||||
Db 25 AGAAAGCAGCAGCAGCC 9
||||| ||| ||||| |||||
RESULT 10
AX476368/c
LOCUS AX476368 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1589 from Patent WO0224750.
ACCESSION AX476368
VERSION AX476368.1 GI:22215653
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1589 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source Location/Qualifiers
1..25
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 0 a 9 c 6 g 10 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
||||| ||| ||||| |||||
Db 24 AGAAAGCAGCAGCAGCC 8
||||| ||| ||||| |||||
RESULT 11
AX476369/c
LOCUS AX476369 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1590 from Patent WO0224750.
ACCESSION AX476369
VERSION AX476369.1 GI:22215654
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1590 28-MAR-2002;
Aeomica, Inc. (US)
FEATURES
source Location/Qualifiers
1..25
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 0 a 9 c 6 g 10 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 8 AGAAAGCAGCAGCAAGCC 24
||||| ||| ||||| |||||
Db 25 AGAAAGCAGCAGCAGCC 9
||||| ||| ||||| |||||

```

      Aecomica, Inc. (US)
FEATURES             Location/Qualifiers
     source             1..25
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT           0 a 9 c 6 g 10 t
ORIGIN

Query Match
Best Local Similarity 57.5%; Score 13.8; DB 6; Length 25;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db 23 AGAAGCACAGCAAGCC 7

RESULT 11
AX476369/c          25 bp DNA linear PAT 12-AUG-2002
LOCUS
DEFINITION          Sequence 1590 from Patent WO0224750.
ACCESSION            AX476369
VERSION              AX476369.1 GI:22215654
KEYWORDS
SOURCE              human.
ORGANISM             Homo sapiens
                    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS             Zhang, J.
TITLE               Human kidney tumor overexpressed membrane protein 1
JOURNAL             Patent: WO 0224750-A 1590 28-MAR-2002;
                    Aecomica, Inc. (US)
FEATURES
     source             1..25
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT           0 a 10 c 6 g 9 t
ORIGIN

Query Match
Best Local Similarity 57.5%; Score 13.8; DB 6; Length 25;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db 22 AGAAGCACAGCAAGCC 6

RESULT 12
AX476370/c          25 bp DNA linear PAT 12-AUG-2002
LOCUS
DEFINITION          Sequence 1591 from Patent WO0224750.
ACCESSION            AX476370
VERSION              AX476370.1 GI:22215655
KEYWORDS
SOURCE              human.
ORGANISM             Homo sapiens
                    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS             Zhang, J.
TITLE               Human kidney tumor overexpressed membrane protein 1
JOURNAL             Patent: WO 0224750-A 1591 28-MAR-2002;
                    Aecomica, Inc. (US)
FEATURES
     source             1..25
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT           1 a 9 c 6 g 9 t
ORIGIN

Query Match
Best Local Similarity 57.5%; Score 13.8; DB 6; Length 25;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db 22 AGAAGCACAGCAAGCC 6

RESULT 13
AX476371/c          25 bp DNA linear PAT 12-AUG-2002
LOCUS
DEFINITION          Sequence 1592 from Patent WO0224750.
ACCESSION            AX476371
VERSION              AX476371.1 GI:22215656
KEYWORDS
SOURCE              human.
ORGANISM             Homo sapiens
                    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS             Zhang, J.
TITLE               Human kidney tumor overexpressed membrane protein 1
JOURNAL             Patent: WO 0224750-A 1592 28-MAR-2002;
                    Aecomica, Inc. (US)
FEATURES
     source             1..25
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT           2 a 8 c 6 g 9 t
ORIGIN

Query Match
Best Local Similarity 57.5%; Score 13.8; DB 6; Length 25;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db 20 AGAAGCACAGCAAGCC 4

RESULT 14
AX476372/c          25 bp DNA linear PAT 12-AUG-2002
LOCUS
DEFINITION          Sequence 1593 from Patent WO0224750.
ACCESSION            AX476372
VERSION              AX476372.1 GI:22215657
KEYWORDS
SOURCE              human.
ORGANISM             Homo sapiens
                    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS             Zhang, J.
TITLE               Human kidney tumor overexpressed membrane protein 1
JOURNAL             Patent: WO 0224750-A 1593 28-MAR-2002;
                    Aecomica, Inc. (US)
FEATURES
     source             1..25
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT           2 a 8 c 6 g 9 t
ORIGIN

Query Match
Best Local Similarity 57.5%; Score 13.8; DB 6; Length 25;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 AGAAGCACAGCAAGCC 24
    ||||| ||| |||
Db 19 AGAAGCACAGCAAGCC 3

RESULT 15
```

AX476373/c
LOCUS AX476373 25 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 1594 from Patent WO0224750.
ACCESSION AX476373
VERSION AX476373.1 GI:22215658
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Zhang, J.
TITLE Human kidney tumor overexpressed membrane protein 1
JOURNAL Patent: WO 0224750-A 1594 28-MAR-2002;
Acemica, Inc. (US)
FEATURES
source 1..25
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2 a 8 c 6 g 9 t
ORIGIN
Query Match 57.5% Score 13.8; DB 6; Length 25;
Best Local Similarity 88.2%; Pred. No. 7e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 8 AGAAGCACGACGAGCC 24
||||||| ||| ||||
Db 18 AGAAGCACGAGCCAGCC 2

Search completed: July 17, 2003, 21:43:07
Job time : 1862 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 19:54:17 ; Search time 227 Seconds
(without alignments)
236.097 Million cell updates/sec

Title: US-09-923-684-12
Perfect score: 24
Sequence: 1 gagagcaagaaagcacagcaagcc 24

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 1698378

Minimum DB seq length: 0
Maximum DB seq length: 25

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	24	Human SIM2 antisense oligonucleotide #2.
2	16	66.7	25	22	Human SIM2 antisense oligonucleotide #2.
3	15.2	63.3	21	24	Human SIM2 antisense oligonucleotide #2.
4	15.2	63.3	24	24	Human SIM2 antisense oligonucleotide #2.
5	15.2	63.3	24	24	Human SIM2 antisense oligonucleotide #2.
6	15.2	63.3	24	24	Human SIM2 antisense oligonucleotide #2.
7	15.2	63.3	24	24	Human SIM2 antisense oligonucleotide #2.
8	15	62.5	20	24	Human SIM2 antisense oligonucleotide #2.
9	14.4	60.0	25	20	Human SIM2 antisense oligonucleotide #2.

10	14.4	60.0	25	24	ABK85446	HIV-1 RNA, second
11	14.2	59.2	20	22	AAH21350	NASBA reaction ass
12	13.8	57.5	17	24	ABQ63834	Human KTM1a portl
13	13.8	57.5	25	24	ABQ64811	Human KTM1a portl
14	13.8	57.5	25	24	ABQ64812	Human KTM1a portl
15	13.8	57.5	25	24	ABQ64813	Human KTM1a portl
16	13.8	57.5	25	24	ABQ64814	Human KTM1a portl
17	13.8	57.5	25	24	ABQ64815	Human KTM1a portl
18	13.8	57.5	25	24	ABQ64816	Human KTM1a portl
19	13.8	57.5	25	24	ABQ64817	Human KTM1a portl
20	13.8	57.5	25	24	ABQ64818	Human KTM1a portl
21	13.8	57.5	25	24	ABQ64819	Human KTM1a portl
22	13.6	56.7	22	20	AAH00990	Human KTM1a portl
23	13.4	55.8	22	20	AAH00990	Human KTM1a portl
24	13.4	55.8	22	20	AAH00990	Human KTM1a portl
25	13.2	55.0	22	16	AAH02239	Human KTM1a portl
26	13.2	55.0	22	18	AAH02239	Human KTM1a portl
27	13.2	55.0	23	16	AAH02239	Human KTM1a portl
28	13.2	55.0	23	16	AAH02239	Human KTM1a portl
29	13.2	55.0	23	16	AAH02239	Human KTM1a portl
30	13.2	55.0	23	16	AAH02239	Human KTM1a portl
31	13.2	55.0	23	16	AAH02239	Human KTM1a portl
32	12.8	53.3	17	20	AAH02239	Human KTM1a portl
33	12.8	53.3	17	20	AAH02239	Human KTM1a portl
34	12.8	53.3	17	20	AAH02239	Human KTM1a portl
35	12.8	53.3	17	20	AAH02239	Human KTM1a portl
36	12.8	53.3	17	20	AAH02239	Human KTM1a portl
37	12.8	53.3	17	20	AAH02239	Human KTM1a portl
38	12.8	53.3	17	20	AAH02239	Human KTM1a portl
39	12.8	53.3	17	20	AAH02239	Human KTM1a portl
40	12.8	53.3	17	20	AAH02239	Human KTM1a portl
41	12.8	53.3	17	20	AAH02239	Human KTM1a portl
42	12.8	53.3	17	20	AAH02239	Human KTM1a portl
43	12.8	53.3	17	20	AAH02239	Human KTM1a portl
44	12.6	52.5	21	23	AAH02239	Human KTM1a portl
45	12.6	52.5	22	24	ABL51638	Human mitotic pits

ALIGNMENTS

RESULT 1	
AD30531	
ID	AD30531 standard; DNA; 24 BP.
XX	
AC	AD30531;
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Human SIM2 antisense oligonucleotide #2.
XX	
KW	Human: single minded homologue 2; SIM2; colon cancer; prostate cancer;
KW	pancreas cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200212565-A1.
XX	
PD	14-FEB-2002.
XX	
PF	06-AUG-2001; 2001WO-US24781.
XX	
PR	04-AUG-2000; 2000US-223531P.
PR	22-DEC-2000; 2000US-257965P.
XX	
PA	(UYFL) UNIV FLORIDA ATLANTIC.
XX	
PI	Narayanan R;
XX	
DR	WPI; 2002-227169/28.
XX	
PT	Detecting cancer in tissue sample, involves providing tissue sample, and analyzing tissue sample for presence of single minded homologue 2

PT marker -

PS Claim 30; Page 58; 60pp; English.

XX The invention relates to a method of detecting cancer in a tissue

CC sample. The method involves providing the tissue sample and analysing

CC the tissue sample for the presence of a single minded homologue 2 (SIM2)

CC nucleic acid or protein, where the presence of the SIM2 marker in the

CC tissue sample indicates that the tissue sample contains cancer. The

CC method is useful for identifying compounds that modulate expression of

CC SIM2 gene in a cell. It is useful for detecting and reducing the growth

CC of cancer such as colon, prostate and pancreas cancer in an animal or

CC mammal. The present sequence is human SIM2 antisense oligonucleotide

CC used in the exemplification of the invention.

XX Sequence 24 BP; 11 A; 6 C; 7 G; 0 U; 0 other;

SQ

Query Match 100.0%; Score 24; DB 24; Length 24;

Best Local Similarity 100.0%; Pred. No. 0.68; 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACGAAGAAGACACAGCAAGCC 24

Db 1 GAGACGAAGAAGACACAGCAAGCC 24

|||||

RESULT 2

AAH39991

ID AAH39991 standard; DNA; 25 BP.

XX AC AAH39991;

XX 14-AUG-2001 (first entry)

XX SNP specific SNPE primer SEQ ID 2787.

XX Single nucleotide polymorphism: SNP; single nucleotide primer extension;

KW SNPE; genotyping; agammaglobulinaemia; diabetes insipidus; cancer;

KW Lesch-Nyhan syndrome; muscular dystrophy; familial hypercholesterolaemia;

KW polycystic kidney disease; osteogenesis imperfecta; autoimmune disease;

KW acute intermittent porphyria; rheumatoid arthritis; multiple sclerosis;

KW inflammation; forensic investigation; paternity analysis; primer; ss.

XX Homo sapiens.

OS

PN WO200129262-A2.

XX 26-APR-2001.

XX 13-OCT-2000; 2000WO-US28436.

XX 15-OCT-1999; 99US-0160096.

XX (ORCH-) ORCHID BIOSCIENCES INC.

XX Picoult-Newburg L, Pohl M;

PI

XX WPI; 2001-290930/30.

DR New genotyping oligonucleotide, useful for detecting the presence,

PT absence or identity of single polynucleotide polymorphism in a nucleic

PT acid sample -

XX Claim 1; Page 64; 83pp; English.

PS

XX Sequences AAH37205 - AAH40944 represent PCR primers, single nucleotide

CC primer extension (SNPE) primers, and the sequences of regions flanking

CC sites of single nucleotide polymorphisms SNPs. The present invention

CC includes kits for determining the presence or absence of a SNP, using the

CC oligonucleotides of the invention. The PCR primers are used to amplify a

CC SNP flanking sequence, the SNPE primer is used as a genotyping primer.

CC The oligonucleotides are useful for genotyping a nucleic acid sample by

CC performing a single-nucleotide primer extension reaction. The

CC oligonucleotides are useful for determining the presence, absence or

CC identity of a SNP and for genotyping nucleic acid samples, for e.g. to

CC assess by association analysis the genotype of an individual or group of

CC individuals, having a pathological phenotypic trait suspected of being

CC caused by one or more SNPs. Phenotypic traits include diseases e.g.

CC agammaglobulinaemia, diabetes insipidus, Lesch-Nyhan syndrome, muscular

CC dystrophy, familial hypercholesterolaemia, polycystic kidney disease,

CC osteogenesis imperfecta and acute intermittent porphyria. Phenotypic

CC traits also include symptoms of or susceptibility to multifactorial

CC disease of which a component is or may be genetic such as autoimmune

CC diseases, including, rheumatoid arthritis, multiple sclerosis,

CC inflammation, cancer, nervous system diseases and infection by pathogenic

CC microorganism. The method is also useful in forensic investigations and

CC paternity analysis. The present sequence represents a single nucleotide

CC primer extension (SNPE) primer specific for a human SNP containing DNA

CC sequence.

XX Sequence 25 BP; 10 A; 8 C; 6 G; 1 T; 0 other;

SQ

Query Match 66.7%; Score 16; DB 22; Length 25;

Best Local Similarity 79.2%; Pred. No. 1.4e+03;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAGACGAAGAAGACACAGCAAGCC 24

Db 1 GAGACGAAGAAGACACATGCC 24

|||||

RESULT 3

ABN81518

ID ABN81518 standard; DNA; 21 BP.

XX AC ABN81518;

XX 13-AUG-2002 (first entry)

XX Yeast PCR primer SEQ ID NO 19.

XX Yeast; pharmaceutical; diarrhoea; intestinal infection; Candida;

KW fermented drink; antidiarrhoeic; fungicide; antibacterial;

KW dermatological; gastrointestinal; PCR; primer; ss.

XX Synthetic.

OS

XX WO200242442-A2.

PN 30-MAY-2002.

XX 15-OCT-2001; 2001WO-EP11887.

XX 24-NOV-2000; 2000DE-1058379.

XX (BIOT-) BIOTECON DIAGNOSTICS GMBH.

XX Grabowski R, Braunschweiger M, Gasch A, Berghof K;

PI

XX WPI; 2002-463630/49.

XX New yeast strains characterized by specific band patterns in a

PT polymerase chain reaction, useful e.g. as probiotics or for preparing

PT fermented drinks -

XX Claim 2; Page 8; 29pp; German.

PS

XX The invention relates to yeast strains (A) that produce a specific band

CC pattern, illustrated in the specification, when characterised by a

CC polymerase chain reaction (PCR). (A), optionally in lyophilised form or

CC as extracts or culture supernatants, are useful for administration to

CC humans or animals, as pharmaceuticals (for treating diarrhoea, colitis,

CC intestinal infections, Candida infections, or skin disorders) or

CC probiotics, also for preparation of fermented drinks, suspensions,

CC extracts and baked goods. (A) have only a minimal effect on the taste of

CC goods prepared using them and can be unequivocally identified by

CC genetic characterisation, even though they are nearly impossible to
CC differentiate biochemically. The present sequence is that of a PCR primer
CC of the invention.

XX
XX
SQ Sequence 21 BP; 7 A; 9 C; 4 G; 1 T; 0 other;

Query Match 63.3%; Score 15.2; DB 24; Length 21;
Best Local Similarity 85.0%; Pred. No. 2.9e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCAGAGAAAGCACAGCAGCC 24
IIIIIIII IIIIIIIII
Db 2 GCAGAGAAATCCAGCCGCAAGCC 21

RESULT 4

ABI82742/C
ID ABI82742 standard; DNA; 24 BP.

XX
AC ABI82742;

XX
DT 15-FEB-2002 (first entry)

XX
DE Capture oligonucleotide zip ID#173 oligo #1.

XX Human; K-ras; PCR primer; probe; capture probe; mutation detection;
KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;
KW infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;
KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;
KW environmental monitoring; food industry; feed industry; ss.

XX Synthetic.

XX OS

XX PN WO200179548-A2.

XX XX

XX PD 25-OCT-2001.

XX PF

XX PR 04-APR-2001; 2001WO-US10958.

XX PR 14-APR-2000; 2000US-197271P.

XX XX (CORR) CORNELL RES FOUND INC.

XX PA Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

XX PI WPI; 2002-034366/04.

XX DR

XX XX

XX PT

XX PT

XX XX

XX PS

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

CC presence or absence of the target nucleotide sequences. ABI82074 to
CC ABI97546 represent oligonucleotide sequences used in the exemplification
CC of the present invention.

XX
XX
SQ Sequence 24 BP; 1 A; 9 C; 4 G; 10 T; 0 other;

Query Match 63.3%; Score 15.2; DB 24; Length 24;
Best Local Similarity 85.0%; Pred. No. 2.9e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 AGCAAGAAAGCACAGCAAGC 23
IIIIIIIIIIIIIIIIIIII
Db 23 AGGGGAAAGCACAGCAAGC 4

RESULT 5

ABI82743

ID ABI82743 standard; DNA; 24 BP.

XX
AC ABI82743;

XX
DT 15-FEB-2002 (first entry)

XX
DE Capture oligonucleotide zip ID#173 oligo #2.

XX Human; K-ras; PCR primer; probe; capture probe; mutation detection;
KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;
KW infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;
KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;
KW environmental monitoring; food industry; feed industry; ss.

XX Synthetic.

XX OS

XX PN WO200179548-A2.

XX XX

XX PD 25-OCT-2001.

XX PF

XX PR 04-APR-2001; 2001WO-US10958.

XX PR 14-APR-2000; 2000US-197271P.

XX XX (CORR) CORNELL RES FOUND INC.

XX PA Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

XX PI WPI; 2002-034366/04.

XX DR

XX XX

XX PT

XX PT

XX XX

XX PS

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

Mon Jul 21 15:46:53 2003

CC presence or absence of the target nucleotide sequences. ABI82074 to
 CC ABI97546 represent oligonucleotide sequences used in the exemplification
 CC of the present invention.

XX Sequence 24 BP; 10 A; 4 C; 9 G; 1 T; 0 other;
 SQ Query Match 63.3%; Score 15.2; DB 24; Length 24;
 Best Local Similarity 85.0%; Pred. No. 2.9e+03;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AGCAAGAAAGCAGCAAGC 23
 || |||||
 Db 2 AGGGGAAAGCAGCAAGC 21

RESULT 6
 ABI92008/c
 ID ABI92008 standard; DNA; 24 BP.

XX ABI92008;
 XX
 DT 15-FEB-2002 (first entry)

Capture oligonucleotide Zip ID#173 oligo #3.

Human; K-ras; PCR primer; probe; capture probe; mutation detection;
 KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;
 KW infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;
 KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;
 KW environmental monitoring; food industry; feed industry; ss.

XX Synthetic.

XX WO200179548-A2.

XX 25-OCT-2001.

XX 04-APR-2001; 2001WO-US10958.

XX 14-APR-2000; 2000US-197271P.

XX (CORR) CORNELL RES FOUND INC.

XX Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

XX WPI; 2002-034366/04.

XX Designing capture oligonucleotide probes for use on a support to which
 XX complementary oligonucleotides hybridize with little mismatch -

XX Claim 3; Fig 26; 300pp; English.

XX The present invention describes a method (M1) for designing capture
 CC oligonucleotide probes (I) for use on a support to which complementary
 CC oligonucleotide probes (II) will hybridize with little mismatch, where
 CC (I) have melting temperatures within a narrow range. The method is useful
 CC for detecting infectious diseases caused by bacterial infectious agents
 CC e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal
 CC infectious agents e.g. Cryptococcus neoformans, Candida albicans and
 CC Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,
 CC Epstein-Barr virus and polio virus, and parasitic infectious agents
 CC selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus
 CC medinis. The method is also useful for detecting genetic diseases such
 CC as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.
 CC Detecting cancer involving oncogenes, tumour suppressor genes, or genes
 CC involved in DNA amplification, replication, recombination or repair, the
 CC cancer is specifically associated with a gene selected from BRCA1 gene,
 CC p53 gene, human papillomavirus types 16 and 18 and liver cancers. The
 CC method is also used for environmental monitoring, forensics and the food
 CC and feed industry, detecting comprises scanning (using e.g. a scanning
 CC electron microscope and infrared microscope) the support at the
 CC particular sites and identifying if ligation of the oligonucleotide probe
 CC sets occurred and correlating (using a computer) identified ligation to a

CC presence or absence of the target nucleotide sequences. ABI82074 to
 CC ABI97546 represent oligonucleotide sequences used in the exemplification
 CC of the present invention.

XX Sequence 24 BP; 1 A; 9 C; 4 G; 10 T; 0 other;
 SQ Query Match 63.3%; Score 15.2; DB 24; Length 24;
 Best Local Similarity 85.0%; Pred. No. 2.9e+03;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AGCAAGAAAGCAGCAAGC 23
 || |||||
 Db 23 AGGGGAAAGCAGCAAGC 4

RESULT 7
 ABI92009
 ID ABI92009 standard; DNA; 24 BP.

XX ABI92009;

XX 15-FEB-2002 (first entry)

XX Capture oligonucleotide Zip ID#173 oligo #4.

Human; K-ras; PCR primer; probe; capture probe; mutation detection;
 KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;
 KW infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;
 KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;
 KW environmental monitoring; food industry; feed industry; ss.

XX Synthetic.

XX WO200179548-A2.

XX 25-OCT-2001.

XX 04-APR-2001; 2001WO-US10958.

XX 14-APR-2000; 2000US-197271P.

XX (CORR) CORNELL RES FOUND INC.

XX Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

XX WPI; 2002-034366/04.

XX Designing capture oligonucleotide probes for use on a support to which
 XX complementary oligonucleotides hybridize with little mismatch -

XX Claim 3; Fig 26; 300pp; English.

XX The present invention describes a method (M1) for designing capture
 CC oligonucleotide probes (I) for use on a support to which complementary
 CC oligonucleotide probes (II) will hybridize with little mismatch, where
 CC (I) have melting temperatures within a narrow range. The method is useful
 CC for detecting infectious diseases caused by bacterial infectious agents
 CC e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal
 CC infectious agents e.g. Cryptococcus neoformans, Candida albicans and
 CC Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,
 CC Epstein-Barr virus and polio virus, and parasitic infectious agents
 CC selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus
 CC medinis. The method is also useful for detecting genetic diseases such
 CC as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.
 CC Detecting cancer involving oncogenes, tumour suppressor genes, or genes
 CC involved in DNA amplification, replication, recombination or repair, the
 CC cancer is specifically associated with a gene selected from BRCA1 gene,
 CC p53 gene, human papillomavirus types 16 and 18 and liver cancers. The
 CC method is also used for environmental monitoring, forensics and the food
 CC and feed industry, detecting comprises scanning (using e.g. a scanning
 CC electron microscope and infrared microscope) the support at the
 CC particular sites and identifying if ligation of the oligonucleotide probe
 CC sets occurred and correlating (using a computer) identified ligation to a

CC presence or absence of the target nucleotide sequences. ABI82074 to
CC ABI97546 represent oligonucleotide sequences used in the exemplification
CC of the present invention.
XX
SQ Sequence 24 BP; 10 A; 4 C; 9 G; 1 T; 0 other;

Query Match 63.3%; Score 15.2; DB 24; Length 24;
Best Local Similarity 85.0%; Pred. No. 2.9e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 AGCAAGAAAGACACACAGC 23
|| |||||
Db 2 AGGGGAAAGACACACAGC 21

RESULT 8

ABI93086/C
ID ABI93086 standard; DNA; 20 BP.

AC ABI93086;

DT 15-FEB-2002 (first entry)

DE Capture oligonucleotide zip ID#173 oligo #9.

KW Human: K-ras; PCR primer; probe; capture probe; mutation detection;
KW ligase detection reaction; LDR; p53; BRCA1; BRCA2; infectious disease;
KW infection; 21 hydroxylase deficiency; Turner Syndrome; obesity;
KW cancer; oncogene; tumour suppressor; human papillomavirus; forensic;
KW environmental monitoring; food industry; feed industry; ss.

OS Synthetic.

PN WO200179548-A2.

PD 25-OCT-2001.

PF 04-APR-2001; 2001WO-US10958.

PR 14-APR-2000; 2000US-197271P.

PA (CORR) CORNELL RES FOUND INC.

PI Barany F, Zirvi M, Gerry NP, Favis R, Kliman R;

DR WPI; 2002-034366/04.

PT Designing capture oligonucleotide probes for use on a support to which
PT complementary oligonucleotides hybridize with little mismatch -

PS Example 5; Fig 29; 300pp; English.

XX The present invention describes a method (M1) for designing capture
XX oligonucleotide probes (I) for use on a support to which complementary
XX oligonucleotide probes (II) will hybridize with little mismatch, where
XX (I) have melting temperatures within a narrow range. The method is useful
XX for detecting infectious diseases caused by bacterial infectious agents
XX e.g. Salmonella, Listeria monocytogenes and Haemophilus influenza, fungal
XX infectious agents e.g. Cryptococcus neoformans, Candida albicans and
XX Aspergillus fumigatus, viruses e.g. T-cell lymphocytotropic virus,
XX Epstein-Barr virus and polio virus, and parasitic infectious agents
XX selected from Onchocerca volvulus, Entamoeba histolytica and Dracunculus
XX medinensis. The method is also useful for detecting genetic diseases such
XX as 21 hydroxylase deficiency, Turner Syndrome and obesity defects.
XX Detecting cancer involving oncogenes, tumour suppressor genes, or genes
XX involved in DNA amplification, replication, recombination or repair, the
XX cancer is specifically associated with a gene selected from BRCA1 gene,
XX p53 gene, human papillomavirus types 16 and 18 and liver cancers. The
XX method is also used for environmental monitoring, forensics and the food
XX and feed industry, detecting comprises scanning (using e.g. a scanning
XX electron microscope and infrared microscope) the support at the
XX particular sites and identifying if ligation of the oligonucleotide probe
XX sets occurred and correlating (using a computer) identified ligation to a

CC presence or absence of the target nucleotide sequences. ABI82074 to
CC ABI97546 represent oligonucleotide sequences used in the exemplification
CC of the present invention.
XX
SQ Sequence 20 BP; 0 A; 7 C; 4 G; 9 T; 0 other;

Query Match 62.5%; Score 15; DB 24; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.5e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GAAGACACACAGC 23
|||||
Db 18 GAAGACACACAGC 4

RESULT 9

AAAX58973/C

ID AAAX58973 standard; DNA; 25 BP.

XX AAAX58973;

AC 23-AUG-1999 (first entry)

DE Human BUB1A kinase 5'RACE primer TJY 300.

KW BUB1A; human; protein kinase; checkpoint control; mitosis;
KW kinetochore; cell proliferation; cancer; tumour; therapy; PCR;
KW RACE; primer; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9928334-A1.

PD 10-JUN-1999.

PF 01-DEC-1998; 98WO-US25415.

PR 01-DEC-1997; 97US-0067093.

PA (FOXC-) FOX CHASE CANCER CENT.

PI Chan G, Jablonski S, Yen T;

DR WPI; 1999-371094/31.

PT Human BUB genes and proteins involved in mitotic checkpoint control

PS Disclosure; Page 31; 99pp; English.

XX This primer oligonucleotide, termed TJY 300, was used with nested
XX 5' Clontech primers AP1 and AP2 to isolate full-length human BUB1A
XX cDNA (see AAAX58970) by extending the yeast two-hybrid Int-57 clone
XX towards the 5' end by RACE. Human BUB1A (see AAY06286) is a novel
XX kinetochore kinase that is involved in mitotic checkpoint control.
XX BUB genes (see AAX58970-72) and their encoded proteins (see AAY06286-88)
XX disclosed in the invention provide valuable therapeutic targets for
XX the design of antiproliferative agents which inhibit the aberrant
XX cellular proliferation observed in tumour cells. Antibodies and
XX oligonucleotide probes based on BUB are also useful in screening
XX for therapeutics, diagnosis and identification of homologues.

SQ Sequence 25 BP; 2 A; 8 C; 4 G; 11 T; 0 other;

Query Match 60.0%; Score 14.4; DB 20; Length 25;
Best Local Similarity 93.8%; Pred. No. 6.3e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GAGCAAGAAAGACAG 18
|||||
Db 25 GAGCAAGAAAGACAG 10

Mon Jul 21 15:46:53 2003

```
RESULT 10
ABK85446
ID ABR85446 standard; DNA; 25 BP.
XX
AC ABR85446;
XX
DT 14-AUG-2002 (first entry)
XX
DE HIV-1 RNA, second primer #6.
XX
KW Human immunodeficiency virus type 1; HIV-1 detection method;
KW primer; ss.
XX
OS Human immunodeficiency virus type 1.
XX
PN EP1203826-A2.
XX
PD 08-MAY-2002.
XX
PF 30-OCT-2001; 2001EP-0125378.
XX
PR 30-OCT-2000; 2000JP-0334937.
XX
PA (TOYO) TOSOH CORP.
XX
PI Ishizuka T, Ishiguro T, Saitoh J;
XX
DR WPI; 2002-473032/51.
XX
XX
XX An oligonucleotide useful for detection of an RNA derived from HIV-1 in
PT clinical tests and diagnosis -
PT
XX Claim 4; Page 21; 34pp; English.
XX
XX The present invention relates to oligonucleotides binding to specific
CC sites of human immunodeficiency virus type 1 (HIV-1) RNA. The
CC oligonucleotides are useful for detecting HIV-1 in clinical tests
CC and diagnosis. The oligonucleotides provide simple, speedy and
CC sensitive detection of HIV-1 RNA which can bind to an intramolecularly
CC free region of the genomic RNA of HIV-1 at relatively low and constant
CC temperatures. The detection method comprises synthesizing a cDNA by
CC the action of an RNA-dependent DNA polymerase by using a specific
CC sequence in an RNA derived from HIV-1 anticipated in a sample as a
CC template, a first primer containing a sequence complementary to the
CC specific sequence and a second primer containing a sequence homologous
CC to the specific sequence (either of which additionally has a promoter
CC sequence for the RNA polymerase at the 5' end). ABR85441-ABK85449
XX represent second primers for HIV-1 RNA.
XX
SQ Sequence 25 BP; 12 A; 6 C; 6 G; 1 T; 0 other;
Query Match 60.0%; Score 14.4; DB 24; Length 25;
Best Local Similarity 93.8%; Pred. No. 6.3e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 AGAAGCACAGCAAGC 23
DB 1 AAAAAGCACAGCAAGC 16
1 |||||

RESULT 11
AAH21350
ID AAH21350 standard; DNA; 20 BP.
XX
AC AAH21350;
XX
DT 17-SEP-2001 (first entry)
XX
DE NASBA reaction associated primer SEQ ID 4.
XX
KW Primer; primer-dependent nucleic acid synthesis; amplification;
KW sequencing; cDNA synthesis; primer-based mutagenesis; ss.
XX
```

```
OS Unidentified.
XX
PN WO200149880-A2.
XX
PD 12-JUL-2001.
XX
PF 27-DEC-2000; 2000WO-EP13288.
XX
PR 30-DEC-1999; 99DE-1063857.
XX
PA (QIAG-) QIAGEN GMBH.
XX
PI Korfhage C, Oelmueller U;
XX
DR WPI; 2001-441896/47.
XX
XX New primers, useful for nucleic acid extension and amplification
PT reactions, with specified base contents in different regions, have
PT increased selectivity and efficiency -
PT
XX Example 1; Page 28; 50pp; German.
XX
XX This invention describes a novel primer (I) in which the proportion of
CC bases that form two hydrogen bonds with the complementary target is at
CC least 50% within the last 6 bases before the 3'-end of the hybridizing
CC portion of (I) and the proportion of bases that form, similarly, three
CC hydrogen bonds is at least 60% within the last 1/4 from the 5'-end of
CC the hybridizing portion. (I), or their mixtures, are used for
CC primer-dependent nucleic acid synthesis and amplification, particularly
CC under isothermal conditions. Typical applications are in sequencing,
CC cDNA synthesis and primer-based mutagenesis. Compared with conventional
CC primers, (I) hybridize very specifically and efficiently to targets,
CC resulting in significantly greater yields with reduced formation of
CC by-products, and eliminate the need to compensate for poor specificity
CC by increasing the temperature. This sequence represents a NASBA nucleic
CC acid sequence based amplification primer used to illustrate the method
CC of the invention.
XX
SQ Sequence 20 BP; 7 A; 4 C; 8 G; 1 T; 0 other;
Query Match 59.2%; Score 14.2; DB 22; Length 20;
Best Local Similarity 84.2%; Pred. No. 7.5e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GAGCAAGAAAGCACAGCAA 21
DB 1 GAGCGGGAAGGCACAGCAA 19
||||| ||| ||||| |||||

RESULT 12
ABQ63834/c
ID ABQ63834 standard; DNA; 17 BP.
XX
AC ABQ63834;
XX
DT 20-AUG-2002 (first entry)
XX
DE Human KTOM1a portion (ABQ63232) probe # 547.
XX
XX Human; KTOM1a; kidney tumour overexpressed membrane; cytostatic;
KW gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
KW kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200224750-A2.
XX
PD 28-MAR-2002.
XX
PF 21-SEP-2001; 2001WO-US29656.
XX
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P;
```

```

PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US00661.
PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 23-MAY-2001; 2001US-0864761.
PR 28-AUG-2001; 2001US-315676P.
XX
XX (AEOM-) AEOMICA INC.
XX
XX Zhang J;
XX
XX WPI; 2002-479509/51.
XX
XX New human kidney tumor overexpressed membrane (KTOM1) protein and
XX nucleic acids encoding the protein, useful for treating subjects having
XX defects in KTOM1 which can manifest as cancer of the kidney, or as a
XX disorder of e.g., liver or bone
XX
XX Example 2; Page 229; 418pp; English.
XX
XX The invention relates to a novel isolated nucleic acid encoding human
XX KTOM1 (kidney tumour overexpressed membrane) protein. The protein of the
XX invention has cytostatic activity. The nucleotide may have a use in gene
XX therapy. The KTOM1 nucleic acids may be used to diagnose, treat or
XX monitor a disease caused by altered expression of human KTOM1.
XX Compositions comprising the nucleic acids, proteins or antibodies may be
XX used to treat subjects having defects in KTOM1 which can manifest as
XX cancer of the kidney, as well as a disorder of liver, bone marrow, brain,
XX heart, lung, kidney, colon, skeletal muscle, testis, uterus and placenta
XX function. The sequence represents a probe used in the invention to
XX scan the nt 1-1001 portion of human KTOM1a (ABQ63232).
XX
XX Sequence 17 BP; 0 A; 5 C; 5 G; 7 T; 0 other;
XX
XX Query Match 57.5%; Score 13.8; DB 24; Length 17;
XX Best Local Similarity 88.2%; Pred. NO. 1.1e+04;
XX Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 8 AGAAAGCAGCAGCAGCC 24
XX | | | | | | | | | | | | | | | |
XX DB 17 AGAAAGCAGCAGCAGCC 1
XX
XX RESULT 13
XX ABQ64811/c
XX ID ABQ64811 standard; DNA; 25 BP.
XX
XX AC ABQ64811;
XX
XX DT 20-AUG-2002 (first entry)
XX
XX DE Human KTOM1a portion (ABQ63232) probe # 1524.
XX
XX KW Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX KW kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200224750-A2.
XX
XX PD 28-MAR-2002.
XX
XX PF 21-SEP-2001; 2001WO-US29656.
XX
XX PR 21-SEP-2000; 2000US-234687P.
XX

```

```

PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US00661.
PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 23-MAY-2001; 2001US-0864761.
PR 28-AUG-2001; 2001US-315676P.
XX
XX (AEOM-) AEOMICA INC.
XX
XX Zhang J;
XX
XX WPI; 2002-479509/51.
XX
XX New human kidney tumor overexpressed membrane (KTOM1) protein and
XX nucleic acids encoding the protein, useful for treating subjects having
XX defects in KTOM1 which can manifest as cancer of the kidney, or as a
XX disorder of e.g., liver or bone
XX
XX Example 2; Page 357; 418pp; English.
XX
XX The invention relates to a novel isolated nucleic acid encoding human
XX KTOM1 (kidney tumour overexpressed membrane) protein. The protein of the
XX invention has cytostatic activity. The nucleotide may have a use in gene
XX therapy. The KTOM1 nucleic acids may be used to diagnose, treat or
XX monitor a disease caused by altered expression of human KTOM1.
XX Compositions comprising the nucleic acids, proteins or antibodies may be
XX used to treat subjects having defects in KTOM1 which can manifest as
XX cancer of the kidney, as well as a disorder of liver, bone marrow, brain,
XX heart, lung, kidney, colon, skeletal muscle, testis, uterus and placenta
XX function. The sequence represents a probe used in the invention to
XX scan the nt 1-1001 portion of human KTOM1a (ABQ63232).
XX
XX Sequence 25 BP; 0 A; 10 C; 5 G; 10 T; 0 other;
XX
XX Query Match 57.5%; Score 13.8; DB 24; Length 25;
XX Best Local Similarity 88.2%; Pred. NO. 1.1e+04;
XX Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 8 AGAAAGCAGCAGCAGCC 24
XX | | | | | | | | | | | | | | | |
XX DB 25 AGAAAGCAGCAGCAGCC 9
XX
XX RESULT 14
XX ABQ64812/c
XX ID ABQ64812 standard; DNA; 25 BP.
XX
XX AC ABQ64812;
XX
XX DT 20-AUG-2002 (first entry)
XX
XX DE Human KTOM1a portion (ABQ63232) probe # 1525.
XX
XX KW Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX KW kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200224750-A2.
XX
XX PD 28-MAR-2002.
XX
XX PF 21-SEP-2001; 2001WO-US29656.
XX
XX PR 21-SEP-2000; 2000US-234687P.
XX

```

```
XX 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US00661.
PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 30-JAN-2001; 2001WO-US00670.
PR 23-MAY-2001; 2001US-0864761.
PR 28-AUG-2001; 2001US-315676P.
XX (AEOM-) AEOMICA INC.
XX Zhang J;
XX WPI; 2002-479509/51.
XX New human kidney tumor overexpressed membrane (KTOM1) protein and
PT nucleic acids encoding the protein, useful for treating subjects having
PT defects in KTOM1 which can manifest as cancer of the kidney, or as a
PT disorder of e.g., liver or bone
XX Example 2; Page 357; 418pp; English.
XX The invention relates to a novel isolated nucleic acid encoding human
CC KTOM1 (kidney tumour overexpressed membrane) protein. The protein of the
CC invention has cytostatic activity. The nucleotide may have a use in gene
CC therapy. The KTOM1 nucleic acids may be used to diagnose, treat or
CC monitor a disease caused by altered expression of human KTOM1.
CC Compositions comprising the nucleic acids, proteins or antibodies may be
CC used to treat subjects having defects in KTOM1 which can manifest as
CC cancer of the kidney, as well as a disorder of liver, bone marrow, brain,
CC heart, lung, kidney, colon, skeletal muscle, testis, uterus and placenta
CC function. The sequence represents a probe used in the invention to
CC scan the nt 1-1001 portion of human KTOM1a (ABQ63232).
XX SQ Sequence 25 BP; 0 A; 9 C; 6 G; 10 T; 0 other;
Query Match 57.5%; Score 13.8; DB 24; Length 25;
Best Local Similarity 88.2%; Pred. NO. 1.1e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 8 AGAAAGCAGCAGCAGCC 24
Db 23 AGAAAGCAGCAGCAGCC 7
RESULT 15
ABQ64813/c
ID ABQ64813 standard; DNA; 25 BP.
XX AC ABQ64813;
XX 20-AUG-2002 (first entry)
XX Human KTOM1a portion (ABQ63232) probe # 1526.
XX Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX Homo sapiens.
XX WO200224750-A2.
XX 28-MAR-2002.
XX 21-SEP-2001; 2001WO-US29656.
```

```
XX 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US00661.
PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 30-JAN-2001; 2001WO-US00670.
PR 23-MAY-2001; 2001US-0864761.
PR 28-AUG-2001; 2001US-315676P.
XX (AEOM-) AEOMICA INC.
XX Zhang J;
XX WPI; 2002-479509/51.
XX New human kidney tumor overexpressed membrane (KTOM1) protein and
PT nucleic acids encoding the protein, useful for treating subjects having
PT defects in KTOM1 which can manifest as cancer of the kidney, or as a
PT disorder of e.g., liver or bone
XX Example 2; Page 357; 418pp; English.
XX The invention relates to a novel isolated nucleic acid encoding human
CC KTOM1 (kidney tumour overexpressed membrane) protein. The protein of the
CC invention has cytostatic activity. The nucleotide may have a use in gene
CC therapy. The KTOM1 nucleic acids may be used to diagnose, treat or
CC monitor a disease caused by altered expression of human KTOM1.
CC Compositions comprising the nucleic acids, proteins or antibodies may be
CC used to treat subjects having defects in KTOM1 which can manifest as
CC cancer of the kidney, as well as a disorder of liver, bone marrow, brain,
CC heart, lung, kidney, colon, skeletal muscle, testis, uterus and placenta
CC function. The sequence represents a probe used in the invention to
CC scan the nt 1-1001 portion of human KTOM1a (ABQ63232).
XX SQ Sequence 25 BP; 0 A; 9 C; 6 G; 10 T; 0 other;
Query Match 57.5%; Score 13.8; DB 24; Length 25;
Best Local Similarity 88.2%; Pred. NO. 1.1e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 8 AGAAAGCAGCAGCAGCC 24
Db 24 AGAAAGCAGCAGCAGCC 8
RESULT 15
ABQ64813/c
ID ABQ64813 standard; DNA; 25 BP.
XX AC ABQ64813;
XX 20-AUG-2002 (first entry)
XX Human KTOM1a portion (ABQ63232) probe # 1526.
XX Human; KTOM1a; KTOM1; kidney tumour overexpressed membrane; cytostatic;
XX gene therapy; cancer; kidney; liver; bone marrow; brain; heart; lung;
XX kidney; colon; skeletal muscle; testis; uterus; placenta; probe; ss.
XX Homo sapiens.
XX WO200224750-A2.
XX 28-MAR-2002.
XX 21-SEP-2001; 2001WO-US29656.
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:12:03 ; Search time 1638 Seconds
(without alignments)
237.297 Million cell updates/sec

Title: US-09-923-684-12
Perfect score: 24
Sequence: .1 gagagcaagaagcacagcaagcc 24

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 18144

Minimum DB seq length: 0
Maximum DB seq length: 25

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:**
2: em_esthum:**
3: em_estin:**
4: em_estmu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_estc:**
9: gb_est1:**
10: gb_est2:**
11: gb_est3:**
12: gb_est4:**
13: gb_est5:**
14: gb_est6:**
15: em_estfun:**
16: em_estcom:**
17: gb_gss:**
18: em_gss_hum:**
19: em_gss_inv:**
20: em_gss_pln:**
21: em_gss_vrt:**
22: em_gss_fun:**
23: em_gss_mam:**
24: em_gss_mus:**
25: em_gss_other:**
26: em_gss_pro:**
27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	12.2	50.8	23	17	AZ469557
C 2	11.6	48.3	23	17	AZ434508
C 3	11.4	47.5	22	17	AZ937666
C 4	11.4	47.5	22	17	TA372612P
C 5	11.4	47.5	23	17	AZ486363
C 6	11.4	47.5	25	17	AZ648037

C 7	11.2	46.7	20	17	AZ453505
C 8	11.2	46.7	23	17	AZ357282
C 9	11.2	46.7	24	17	AZ999558
C 10	11	45.8	19	17	AZ774205
C 11	11	45.8	20	17	AZ834080
C 12	11	45.8	24	17	AZ778465
C 13	10.8	45.0	20	17	AZ796553
C 14	10.8	45.0	20	17	AZ608800
C 15	10.6	44.2	21	17	AZ308814
C 16	10.6	44.2	22	9	AI808665
C 17	10.6	44.2	22	17	AZ443174
C 18	10.6	44.2	23	17	AZ318263
C 19	10.4	43.3	21	17	AZ789335
C 20	10.4	43.3	24	17	AZ402832
C 21	10.4	43.3	24	17	AZ588136
C 22	10.4	43.3	24	17	AZ815479
C 23	10.2	42.5	21	17	AZ656028
C 24	10.2	42.5	22	17	AZ941882
C 25	10.2	42.5	23	17	AZ973926
C 26	10.2	42.5	24	17	AZ307138
C 27	10.2	42.5	24	17	AZ792444
C 28	10.2	42.5	25	9	AI444500
C 29	10.2	42.5	25	9	AA469268
C 30	10.2	42.5	25	17	AZ389918
C 31	10	41.7	21	17	AZ334510
C 32	10	41.7	21	17	AZ336281
C 33	10	41.7	22	17	AZ471500
C 34	10	41.7	23	17	AZ466178
C 35	10	41.7	25	9	AA988825
C 36	10	41.7	25	17	AZ465511
C 37	10	41.7	25	17	AZ810763
C 38	9.8	40.8	19	17	AZ406137
C 39	9.8	40.8	21	17	AZ308773
C 40	9.8	40.8	21	17	AZ346717
C 41	9.8	40.8	22	9	AI344786
C 42	9.8	40.8	22	17	AZ307823
C 43	9.8	40.8	22	17	AZ324102
C 44	9.8	40.8	22	17	AZ950407
C 45	9.8	40.8	23	17	AZ430288

ALIGNMENTS

RESULT 1
AZ469557/c

LOCUS
DEFINITION

ACCESSION
VERSION

KEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AZ469557 23 bp DNA linear GSS 04-OCT-2000
IM0283A09F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
clone UUGCIM0283A09 F, DNA sequence.

AZ469557
GSS.
GI:10627682

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 23)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0220 row: K column: 22
Seq primer: CACACAGAAACAGCATGACC
Class: plasmid ends
High quality sequence stop: 23.
Location/Qualifiers
1. .23
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0220K22"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 9 a 8 c 4 g 2 t
ORIGIN
Query Match 48.3%; Score 11.6; DB 17; Length 23;
Best Local Similarity 77.8%; Pred. No. 5.6e+05;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 AAGAAAGCAGCAAGCC 24
||||| ||||| ||||| ||
Db 3 AAGCACCAGCAAGCC 20

RESULT 3
AZ937666/c
LOCUS
DEFINITION
2M0195L23R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0195L23 R, DNA sequence.
ACCESSION
AZ937666
VERSION
AZ937666.1 GI:13796671
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 22)
REFERENCE
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

Plate: 0283 row: A column: 09
Seq primer: CGTTGTAACAGCGGCCAGT
Class: plasmid ends
High quality sequence stop: 23.
Location/Qualifiers
1. .23
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0283A09"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 0 a 9 c 2 g 12 t
ORIGIN
Query Match 50.8%; Score 12.2; DB 17; Length 23;
Best Local Similarity 82.4%; Pred. No. 3.2e+05;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGACGAGAGAGACACA 17
||||| ||||| ||||| ||
Db 19 GAGAGAGAGAGAGACACA 3

RESULT 2
AZ434508
LOCUS
DEFINITION
IM0220K22R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0220K22 R, DNA sequence.
ACCESSION
AZ434508
VERSION
AZ434508.1 GI:10558521
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 23)
REFERENCE
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

Plate: 0195 row: L column: 23
 Seq primer: CACAGGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 22.

FEATURES

Location/Qualifiers
 1. .22
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUC2M0195L23"
 /clone_lib="Mouse 10kb plasmid UUC2M library"
 /sex="Female"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 2 a 11 c 2 g 7 t

ORIGIN

Query Match 47.5%; Score 11.4; DB 17; Length 22;
 Best Local Similarity 92.3%; Pred. No. 6.6e+05;
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGAGCAAGAAG 13

Db 22 GAGAGCAAGAAG 10

RESULT 4

TA372G12P 22 bp DNA linear GSS 13-DEC-2000
 T. brucei sheared genomic DNA clone 372g12, forward sequence,
 genomic survey sequence.

LOCUS AL496138

VERSION AL496138.1 GI:11872177

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 22)

AUTHORS Hall, N., Bowman, S., Lennard, N. J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S. E., Rajandream, M. A. and Barrell, B. G.

TITLE Direct Submission

JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhle@sanger.ac.uk
 Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREP927/4 Gurat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J. C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

FEATURES

source

Location/Qualifiers
 1. .22
 /organism="Trypanosoma brucei"
 /strain="TREP927"
 /db_xref="taxon:5691"
 /clone="372g12"

BASE COUNT 11 a 3 c 8 g 0 t

ORIGIN

Query Match 47.5%; Score 11.4; DB 17; Length 22;
 Best Local Similarity 71.4%; Pred. No. 6.6e+05;
 Matches 15; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 AGAGCAAGAAAGCACAGCAAG 22

Db 2 AGCGAAGGAGCGAAGAAAG 22

RESULT 5

AZ486363/c

LOCUS 1M0314015F

DEFINITION Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0314015 F, DNA sequence.

ACCESSION AZ486363

VERSION AZ486363.1 GI:10653065

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 23)

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D. Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0314 row: 0 column: 15

Seq primer: CGTTGTAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 23.

Location/Qualifiers

1. .23

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0314015"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at http://www.sanger.ac.uk/projects/T-brucei/.

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 5 a 4 c 7 g 7 t
ORIGIN
Query Match 47.5%; Score 11.4; DB 17; Length 23;
Best Local Similarity 92.3%; Pred. No. 6.7e+05;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 AAGCAGCAGCAGC 23
|||||
Db 21 AAGCAGCAGCAGC 9

RESULT 6
AZ648037/c
LOCUS
DEFINITION
1M0514C23R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0514C23 R, DNA sequence.

ACCESSION
AZ648037
VERSION
AZ648037.1 GI:11780102
KEYWORDS
GSS.

SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 25)
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0514 row: C column: 23
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 25.
Location/Qualifiers
1. .25

FEATURES
source
1. .25
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0514C23"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 0 a 6 c 9 g 10 t
ORIGIN
Query Match 47.5%; Score 11.4; DB 17; Length 25;
Best Local Similarity 92.3%; Pred. No. 6.8e+05;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 AGAAGCAGCAGCA 20
|||||
Db 16 AGACAGCAGCAGCA 4

RESULT 7
AZ453505/c
LOCUS
DEFINITION
1M0254M22R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0254M22 R, DNA sequence.

ACCESSION
AZ453505
VERSION
AZ453505.1 GI:10611371
KEYWORDS
GSS.

SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 20)
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0254 row: M column: 22
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 20.
Location/Qualifiers
1. .20

FEATURES
source
1. .20
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0254M22"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gii14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

BASE COUNT      2 a      4 c      2 g      12 t
ORIGIN
Query Match      46.7%; Score 11.2; DB 17; Length 20;
Best Local Similarity 81.2%; Pred. No. 7.7e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy  2 AGAGCAAGAAAGCACA 17
    ||||| |||| |||
Db  19 AGAGCATGAAGAGACA 4

RESULT 8
AZ357282/c
LOCUS      23 bp      DNA      linear      GSS 02-OCT-2000
DEFINITION 1M0098A16R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
ACCESSION  AZ357282
VERSION     AZ357282.1 GI:10470982
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
          Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
          ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
          and Wright,D.,Weiss,R.
          Mouse whole genome scaffolding with paired end reads from 10kb
          plasmid inserts
          Unpublished (2000)
          Contact: Robert B. Weiss
          University of Utah Genome Center
          Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
          84112, USA
          Tel: 801 585 5606
          Fax: 801 585 7177
          Email: ddunn@genetics.utah.edu
          Insert Length: 10000 Std Error: 0.00
          Plate: 0098 row: A column: 16
          Seq primer: CACACAGAAACAGCTATGACC
          Class: plasmid ends
          High quality sequence stop: 23.
          Location/Qualifiers
            1. 23
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="UUGC1M0098A16"
              /clone_lib="Mouse 10kb plasmid UUGC1M library"
              /sex="Male"
              /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              musculus C57BL/6J (male) was obtained from the Jackson
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
  
```

```

BASE COUNT      2 a      4 c      3 g      14 t
ORIGIN
Query Match      46.7%; Score 11.2; DB 17; Length 23;
Best Local Similarity 81.2%; Pred. No. 8e+05;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy  7 AAGAAACACAGCAAG 22
    ||||| |||| |||
Db  16 AAGAAACACACAAAG 1

RESULT 9
AZ999558/c
LOCUS      24 bp      DNA      linear      GSS 27-APR-2001
DEFINITION 2M0286N24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
ACCESSION  AZ999558
VERSION     AZ999558.1 GI:13870785
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
          Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
          ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
          and Wright,D.,Weiss,R.
          Mouse whole genome scaffolding with paired end reads from 10kb
          plasmid inserts
          Unpublished (2000)
          Contact: Robert B. Weiss
          University of Utah Genome Center
          Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
          84112, USA
          Tel: 801 585 5606
          Fax: 801 585 7177
          Email: ddunn@genetics.utah.edu
          Insert Length: 10000 Std Error: 0.00
          Plate: 0286 row: N column: 24
          Seq primer: CACACAGAAACAGCTATGACC
          Class: plasmid ends
          High quality sequence stop: 24.
          Location/Qualifiers
            1. 24
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="UUGC2M0286N24"
              /clone_lib="Mouse 10kb plasmid UUGC2M library"
              /sex="Female"
              /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              musculus C57BL/6J (female) was obtained from the Jackson
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
  
```

```

FEATURES
source
1. 24
  /organism="Mus musculus"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="UUGC2M0286N24"
  /clone_lib="Mouse 10kb plasmid UUGC2M library"
  /sex="Female"
  /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
  /note="Vector: PWD42nv; Purified genomic DNA from M.
  musculus C57BL/6J (female) was obtained from the Jackson
  Laboratory Mouse DNA Resource
  (http://www.jax.org/resources/documents/dnares/). The DNA
  was hydrodynamically sheared by repeated passage through a
  0.005 inch orifice at constant velocity. The sheared DNA
  was blunt end-repaired with T4 DNA polymerase and T4
  polynucleotide kinase. Adaptor oligonucleotides were
  ligated to the blunt ends in high molar excess. The
  
```


adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 1 a 4 c 4 g 11 t
ORIGIN

Query Match 45.8%; Score 11; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAA 12
|||||
Db 11 AGAGCAAGAAA 1

RESULT 12
AZ778465/c

LOCUS 24 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0013G17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0013G17 R, DNA sequence.

ACCESSION AZ778465
VERSION AZ778465.1 GI:12908136
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 24)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
Plate: 0013 row: G column: 17
Seq primer: CACACAGGAACACGTCATGACC
Class: plasmid ends
High quality sequence stop: 24.

FEATURES
source

1. .24
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0013G17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 6 c 5 g 10 t
ORIGIN

Query Match 45.8%; Score 11; DB 17; Length 24;
Best Local Similarity 73.7%; Pred. No. 9.8e+05;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGAGCAAGAACACACGCA 20
|||||
Db 21 AGAGAAAGAGCTCACTGCA 3

RESULT 13
AZ796553

LOCUS 20 bp DNA linear GSS 16-FEB-2001
DEFINITION 2M0052P15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0052P15 F, DNA sequence.

ACCESSION AZ796553
VERSION AZ796553.1 GI:12944728
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 20)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
Plate: 0052 row: P column: 15
Seq primer: CGTGTAAACACGCGCCACT
Class: plasmid ends
High quality sequence stop: 20.
Location/Qualifiers
1. .20
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0052P15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

FEATURES
source

1. .20
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0052P15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 11 a 2 c 6 g 1 t
ORIGIN

Query Match 45.0%; Score 10.8; DB 17; Length 20;
Best Local Similarity 85.7%; Pred. NO. 1.1e+06;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGACAAGAAGCA 15
||||| ||||| |||
Db 6 AGAGAAGAATGCA 19

RESULT 14
AZ808800/c
LOCUS
DEFINITION
2M0072F01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0072F01 R, DNA sequence.

ACCESSION
AZ808800
VERSION
GSS.
KEYWORDS
SOURCE
house mouse.
Mus musculus

ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 20)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.,
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

TITLE
Unpublished (2000)
JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0072 row: F column: 01
Seq primer: CACACAGAACACGATGACC
Class: plasmid ends
High quality sequence stop: 20.
Location/Qualifiers

FEATURES
source
1. .20
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0072F01"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 1 a 8 c 3 g 8 t
ORIGIN

Query Match 45.0%; Score 10.8; DB 17; Length 20;
Best Local Similarity 85.7%; Pred. NO. 1.1e+06;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCAAGAAAGCACAG 18
||||| ||||| |||
Db 20 GCAAGAAAGGAAAG 7

RESULT 15
AZ308814/c
LOCUS
DEFINITION
1M0012N07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0012N07 F, DNA sequence.

ACCESSION
AZ308814
VERSION
GSS.
KEYWORDS
SOURCE
house mouse.
Mus musculus

ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 21)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.,
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

TITLE
Unpublished (2000)
JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0012 row: N column: 07
Seq primer: CGTTGTAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 21.
Location/Qualifiers

FEATURES
source
1. .21
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0012N07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWP42 (g147321419b/Ar129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 8 c 2 g 8 t
ORIGIN

Query Match 44.2%; Score 10.6; DB 17; Length 21;
Best Local Similarity 76.5%; Pred. No. 1.4e+06;
Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACAGCAGCAACACACAG 18
||||| ||| ||||| ||
Db 17 AGAGTAAGTAGCGGAG 1

Search completed: July 17, 2003, 22:19:04
Job time : 1646 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:01:47 : Search time 48 Seconds
(without alignments)
153.338 Million cell updates/sec

Title: US-09-923-684-12
Perfect score: 24
Sequence: 1 gagagcaagaagcacagcagcc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 364038

Minimum DB seq length: 0
Maximum DB seq length: 25

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13.6	56.7	22	4	US-09-467-997-7
2	13.4	55.8	24	2	US-08-117-952-239
3	13.2	55.0	22	2	US-08-634-797-9
4	13.2	55.0	23	1	US-08-406-635-10
5	13	54.2	21	1	US-08-263-413-21
6	12.8	53.3	21	2	US-08-117-952-295
7	12.6	52.5	23	4	US-09-338-907-143
8	12.6	52.5	23	4	US-09-218-207-143
9	12.4	51.7	19	4	US-09-147-208-70
10	12.4	51.7	25	4	US-09-315-794-8
11	12.4	51.7	25	4	US-09-389-341-8
12	12.2	50.8	18	3	US-09-339-964-46
13	12.2	50.8	20	1	US-08-729-447-15
14	12.2	50.8	20	2	US-08-715-890-1
15	12.2	50.8	20	2	US-08-743-6378-159
16	12.2	50.8	20	3	US-08-526-8408-159
17	12.2	50.8	22	1	US-08-729-447-12
18	12.2	50.8	25	1	US-08-410-804-11
19	12.2	50.8	25	1	US-08-607-269-14
20	12.2	50.8	25	1	US-08-259-514-11
21	12.2	50.8	25	2	US-08-858-311-11
22	12.2	50.8	25	4	US-08-697-610-7
23	12.2	50.8	25	4	US-08-349-357-7
24	12.2	50.8	25	5	PCT-US95-04600-14
25	12	50.0	19	1	US-08-268-799-4
26	12	50.0	19	6	5166195-2
27	12	50.0	20	1	US-08-308-869-2

C 28	12	50.0	20	1	US-08-782-980-2	Sequence 2, Appli
C 29	12	50.0	20	2	US-08-808-474A-19	Sequence 19, Appl
C 30	12	50.0	20	2	US-08-808-474A-20	Sequence 20, Appl
C 31	12	50.0	20	4	US-09-235-614-19	Sequence 19, Appl
C 32	12	50.0	20	4	US-09-235-614-20	Sequence 20, Appl
C 33	12	50.0	20	4	US-09-517-584A-63	Sequence 63, Appl
C 34	12	50.0	20	4	US-09-036-637D-2	Sequence 2, Appli
C 35	12	50.0	20	4	US-09-036-637D-5	Sequence 5, Appli
C 36	12	50.0	21	6	5166193-3	Patent No. 5166195
C 37	12	50.0	22	4	US-09-092-077-20	Sequence 20, Appl
C 38	12	50.0	24	1	US-08-412-229-8	Sequence 8, Appli
C 39	12	50.0	24	1	US-08-825-617-8	Sequence 8, Appli
40	11.8	49.2	20	1	US-08-602-203-1	Sequence 1, Appli
41	11.8	49.2	20	3	US-09-009-913-331	Sequence 331, App
C 42	11.8	49.2	20	4	US-09-517-4678-303	Sequence 303, App
C 43	11.8	49.2	23	2	US-08-687-865A-12	Sequence 12, Appl
44	11.8	49.2	23	4	US-09-043-711-12	Sequence 12, Appl
C 45	11.8	49.2	23	4	US-09-240-918-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-467-997-7/c
; Sequence 7, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
; APPLICANT: Kitajewski, Jan
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
; FILE REFERENCE: 53863-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/467,997
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 22
; TYPE: DNA
; ORGANISM: mouse
US-09-467-997-7

Query Match 56.7%; Score 13.6; DB 4; Length 22;
Best Local Similarity 80.0%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCAAGAAAGCACAGCAAGCC 24
DB 21 GCAAGAAAGCGCAGCAGGAC 2

RESULT 2
US-08-117-952-239
; Sequence 239, Application US/08117952
; Patent No. 5851760
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Smith, Michael W.
; TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE
; TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES
; NUMBER OF SEQUENCES: 797
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,952
; FILING DATE: 07-SEP-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/078,471
; FILING DATE: 15-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 239:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Oligonucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-117-952-239

Query Match 55.8%; Score 13.4; DB 2; Length 24;
Best Local Similarity 73.9%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCAGCAAGCC 24
Db 2 ACAGCAAAARACTAAGCAATCC 24

RESULT 3
US-08-634-797-9/c
; Sequence 9, Application US/08634797
; Patent No. 5851759
; GENERAL INFORMATION:
; APPLICANT: WEINER, AMY J.
; TITLE OF INVENTION: HETERO DUPLEX TRACKING ASSAY (HTA) FOR
; TITLE OF INVENTION: GENOTYPING HCV
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street - R440
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/634,797
; FILING DATE: 19-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 1226.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-3274
; TELEFAX: (510) 655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;

;
; MOLECULE TYPE: DNA (genomic)
; US-08-634-797-9

Query Match 55.0%; Score 13.2; DB 2; Length 22;
Best Local Similarity 83.3%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AGCAAGAAAGCAGCAAGCA 21
Db 18 AGCAAGAGACAGAGCA 1

RESULT 4
US-08-406-635-10/c
; Sequence 10, Application US/08406635
; Patent No. 559674
; GENERAL INFORMATION:
; APPLICANT: PENA, SERGIO D.J.
; APPLICANT: SIMPSON, ANDREW J.G.
; TITLE OF INVENTION: METHOD FOR RECOGNITION OF THE NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE OF A PURIFIED DNA SEGMENT
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,635
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/100,738
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 45119-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-406-635-10

Query Match 55.0%; Score 13.2; DB 1; Length 23;
Best Local Similarity 83.3%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 CAAGAAAGCAGCAAGC 23
Db 19 CAAGCAAGTACAGCAATC 2

RESULT 5
US-08-263-413-21
; Sequence 21, Application US/08263413
; Patent No. 5747246
; GENERAL INFORMATION:
; APPLICANT: PANNETIER, CHRISTOPHE
; APPLICANT: COCHET, MADELEINE
;

```


APPLICANT: DANCHE, SYLVIE
APPLICANT: KOURILSKY, PHILIPPE
TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF
TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF
TITLE OF INVENTION: METHOD OF ENZYMIC AMPLIFICATION OF DNA
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/263,413
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,980
FILING DATE: 14-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, J.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 660-058-55X
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-263-413-21

Query Match 54.28; Score 13; DB 1; Length 21;
Best Local Similarity 76.2; Pred. No. 2e+03;
Matches 16; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Qy 4 AGCAAGAAAGCAGCAGG 24
Db 1 AACAAAGAAAGCAGCAGG 21

RESULT '6

US-08-117-952-295/c
Sequence 295, Application US/08117952
Patent No. 5851760
GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
APPLICANT: Smith, Michael W.
TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE
TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES
NUMBER OF SEQUENCES: 797
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,952
FILING DATE: 07-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/078,471
FILING DATE: 15-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9423
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 295:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Oligonucleotide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-117-952-295

Query Match 53.3%; Score 12.8; DB 2; Length 21;
Best Local Similarity 87.5%; Pred. No. 2.5e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GAGCAAGAAAGCAGCAG 18
Db 16 GAACATGAAGCAGCAG 1

RESULT 7

US-09-338-907-143
Sequence 143, Application US/09338907
Patent No. 6265546
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilyia, Chumakov
APPLICANT: Bouqueleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET.18CPICP
CURRENT APPLICATION NUMBER: US/09/338,907
CURRENT FILING DATE: 1999-06-23
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 09/218,207
EARLIER FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 143
LENGTH: 23
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..23
OTHER INFORMATION: amplification oligonucleotide PGIase24
US-09-338-907-143

Query Match 52.5%; Score 12.6; DB 4; Length 23;
Best Local Similarity 78.9%; Pred. No. 3.1e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 AGCAAGAAAGCAGCAGG 22
Db 2 ATCATCAAGCAGCAGCATG 20

```

; LENGTH: 19 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-147-208-70

Query Match          51.7%; Score 12.4; DB 4; Length 19;
Best Local Similarity 92.9%; Pred. No. 3.6e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      9 GAAAGCACGACGACG 22
Db      18 GAAAGCACGACGACG 5

RESULT 10
US-09-315-794-8
; Sequence 8, Application US/09315794
; Patent No. 6197517
; GENERAL INFORMATION:
; APPLICANT: Robert's, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-053
; CURRENT APPLICATION NUMBER: US/09/315,794
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
; US-09-315-794-8

Query Match          51.7%; Score 12.4; DB 4; Length 25;
Best Local Similarity 92.9%; Pred. No. 3.8e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 AGCAAGAAAGCACCA 17
Db      7 AGAAGAAAGCACCA 20

RESULT 11
US-09-389-341-8
; Sequence 8, Application US/09389341
; Patent No. 6200803
; GENERAL INFORMATION:
; APPLICANT: Robert's, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-057
; CURRENT APPLICATION NUMBER: US/09/389,341
; CURRENT FILING DATE: 1999-09-02
; EARLIER APPLICATION NUMBER: 09/315,794
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
; US-09-389-341-8

Query Match          51.7%; Score 12.4; DB 4; Length 25;
Best Local Similarity 92.9%; Pred. No. 3.8e+03;
```

```

; LENGTH: 23
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..23
; OTHER INFORMATION: amplification oligonucleotide PGLASe24
; US-09-218-207-143

Query Match          52.5%; Score 12.6; DB 4; Length 23;
Best Local Similarity 78.9%; Pred. No. 3.1e+03;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 AGCAAGAAAGCACGACGACG 22
Db      2 ATCATCAAGACGACGACG 20

RESULT 9
US-09-147-208-70/c
; Sequence 70, Application US/09147208
; Patent No. 6333303
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Antiviral Ricin-Like Proteins
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/147,208
; APPLICATION NUMBER: US/09/147,208
; FILING DATE: 02-MAR-1999
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Rudolph, John R.
; REGISTRATION NUMBER: 38,003
; REFERENCE/DOCKET NUMBER: 7841-76
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
```

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 AGCAAGAAAGCACCA 17
Db 7 AGAAGAAAGCACCA 20

RESULT 12
US-09-339-964-46
; Sequence 46, Application US/09339964
; Patent No. 6025198
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF SHIP-2 EXPRESSION
; FILE REFERENCE: RTS-0065
; CURRENT APPLICATION NUMBER: US/09/339,964
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 46
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-339-964-46

Query Match 50.8%; Score 12.2; DB 3; Length 18;
Best Local Similarity 82.4%; Pred. No. 4.4e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AGCAAGAAAGCACAG 18
Db 1 AGCAGGAGGACACAG 17

RESULT 13
US-08-729-447-15
; Sequence 15, Application US/08729447
; Patent No. 5789174
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DETECTION OF PERIODONTAL PATHOGENS INCLUDING
; TITLE OF INVENTION: BACTEROIDES FORSYTHUS, PORPHYROMONAS GINGIVALIS,
; TITLE OF INVENTION: PREVOTELLA INTERMEDIA AND PREVOTELLA NIGRESCENS
; NUMBER OF SEQUENCES: 20
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/08/729,447
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "PCR primer pair No. 5789174 3 for
; DESCRIPTION: P. intermedia"
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-729-447-15

Query Match 50.8%; Score 12.2; DB 1; Length 20;
Best Local Similarity 82.4%; Pred. No. 4.5e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AGCAAGAAAGCACCA 20

Db 2 AGCAAGAAAGCAAGGA 18

RESULT 14
US-08-715-890-1
; Sequence 1, Application US/08715890
; Patent No. 5919622
; GENERAL INFORMATION:
; APPLICANT: MACHO, Heinz
; APPLICANT: BEINHAUS, Gerhard
; TITLE OF INVENTION: SYSTEM FOR THE TEMPERATURE
; TITLE OF INVENTION: ADJUSTMENT TREATMENT OF LIQUID SAMPLES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAYDO, MARCELSTEIN, MURRAY & ORAM LLP
; STREET: 655 Fifteenth Street, N. W., Suite 330 - G
; STREET: Street Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/715,890
; FILING DATE: 19-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERMAN, Richard J.
; REGISTRATION NUMBER: 39,107
; REFERENCE/DOCKET NUMBER: PL614-6054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Oligodesoxyribonucleotide"
; HYPOTHETICAL: NO
US-08-715-890-1

Query Match 50.8%; Score 12.2; DB 2; Length 20;
Best Local Similarity 82.4%; Pred. No. 4.5e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 8 AGAAGACAGCAGCAAGCC 24
Db 1 AGACAGTACAGCAGCC 17

RESULT 15
US-08-743-637B-159/c
; Sequence 159, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:

ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 159:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
US-08-743-637B-159

Query Match 50.8%; Score 12.2; DB 2; Length 20;
Best Local Similarity 82.4%; Pred No. 4.5e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGAGCAAGAAAGCACAG 18
||| ||||| |||
Db 19 AGAACAAGAAAGGAAAG 3

Search completed: July 17, 2003, 21:51:26
Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 16:48:04 ; Search time 7266 Seconds
(without alignments)
11451.269 Million cell updates/sec

Title: US-09-923-684-2
Perfect score: 2859
Sequence: 1 actcactataggctcgagc.....gtctcaaaaaaaaaaaaaa 2059
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_rtd.*
- 37: em_htg_vit.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB	ID	Description
1	2859	100.0	2859	9	HSU80457	U80457 Human trans
2	1661.6	58.1	3921	9	HSU80456	U80456 Human trans
3	1593.6	55.7	136219	9	AP000697	AP000697 Homo sapi
4	1593.6	55.7	340000	9	AP001726	AP001726 Homo sapi
5	1590.4	55.6	3365	9	AB003185	AB003185 Homo sapi
6	1584.8	55.4	120000	2	HSAC000001	AC000001 Homo sapi
7	1389	48.6	1465	9	HS001858	AJ001858 Homo sapi
8	1323.4	46.3	3614	6	EI4065	EI4065 cDNA encodi
9	1323.4	46.3	3614	10	D64135	D64135 Mus musculu
10	1321.8	46.2	2914	10	MUSMSIM6J	D63383 Mouse mRNA
11	1320.2	46.2	3071	10	MMU42554	U42554 Mus musculu
12	1257.2	44.0	3963	10	MMU40576	U40576 Mus musculu
13	1142.8	40.0	250000	9	AB063285	AB063285 Homo sapi
14	939	32.8	2309	5	AF283298	AF283298 Xenopus l
15	704.6	24.6	4007	9	HSU70212	U70212 Human SIM1
16	671.8	23.5	2547	10	D79209	D79209 Mus musculu
17	665.8	23.3	2664	5	AY028626	AY028626 Danio rer
18	662.8	23.2	4380	5	AF363019	AF363019 Danio rer
19	655.8	22.9	7369	10	MMU40575	U40575 Mus musculu
20	397.6	13.9	2838	3	AV129457	AV129457 Drosophil
21	370.6	13.0	2747	3	DROSIMNP	MI9020 D.melanogas
22	296	10.4	1532	10	MMSM2G09	AF023872 Mus muscu
23	234	8.2	42338	9	HUMQ16F8	D85922 Human DNA f
24	233.8	8.2	147054	2	AL356582	AL356582 Homo sapi
25	232.6	8.1	160810	9	AC009953	AC009953 Homo sapi
26	232.2	8.1	167548	9	AL391809	AL391809 Human DNA
27	231.6	8.1	162771	9	AC027121	AC027121 Homo sapi
28	231.6	8.1	163157	2	AC012567	AC012567 Homo sapi
29	231.6	8.1	166338	9	AC026165	AC026165 Homo sapi
30	231.6	8.1	208879	9	AC024171	AC024171 Homo sapi
31	231.4	8.1	238	6	EI4059	EI4059 A part of h
32	231.4	8.1	238	9	HUMSIMA	D44444 Homo sapien
33	231.2	8.1	178367	9	AC006115	AC006115 Homo sapi
34	231	8.1	43128	9	AC011558	AC011558 Homo sapi
35	231	8.1	156197	9	HS97K10	281365 Human DNA s
36	230.6	8.1	50000	6	AX393458	AX393458 Sequence
37	230.6	8.1	83549	9	AC004752	AC004752 Homo sapi
38	230.6	8.1	107480	9	AL596094	AL596094 Human DNA
39	230.6	8.1	184268	2	AC011401	AC011401 Homo sapi
40	230.6	8.1	208145	2	AC053504	AC053504 Homo sapi
41	230.4	8.1	3605	5	AF304864	AF304864 Homo sapi
42	230.4	8.1	137074	9	AC072052	AC072052 Homo sapi
43	230.4	8.1	195342	9	AC009120	AC009120 Homo sapi
44	230	8.0	115231	9	HS785G19	AL035458 Human DNA
45	230	8.0	182753	2	AL137253	AL137253 Homo sapi

ALIGNMENTS

RESULT 1
HSU80457
LOCUS
DEFINITION Human transcription factor SIM2 short form mRNA linear PRI 08-JUL-1997
ACCESSION U80457
VERSION U80457.1 GI:2062418
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2859)
AUTHORS Chast,R., Scott,H.S., Chen,H., Kudoh,J., Rossier,C., Minoshima,S.,
Wang,Y., Shimizu,N. and Antonarakis,S.E.
TITLE Cloning of two human homologs of the Drosophila single-minded gene

SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome
chromosomal region
Genome Res. 7 (6), 615-624 (1997)
97343329
PUBMED
9199934
REFERENCE
AUTHORS
2 (bases 1 to 2859)
Chrast,R., Kudoh,J., Rossier,C., Chen,H., Minoshima,S., Shimizu,N.
and Antonarakis,S.E.
Direct Submission
Submitted (29-NOV-1996) Medical Genetics, University of Geneva
Medical School, 1, Rue Michel-Servet, Geneva 1211, Switzerland
Location/Qualifiers
1..2859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.2"
93..1805
/codon_start=1
/product="transcription factor SIM2 short form"
/protein_id="AAB62397.1"
/db_xref="GI:2062419"
/translation="MKESKNAKTRREKNGEYELAKLLPLPSAITSOLDKASIIIR
LITSLKMRVPEGLGDWQPSRAGPLDGVAKELGSHLIQTLDTGDFVVFVSDGKIM
YISFASVHLKLSQVELTGNISYIYIHPSDHDEMTAVLTAHOPHLHLLOEIERSEF
FLRMKCVLAKRNAGLTCGKYIHCISGKLRQYMLDMSLYDSCQIYLVHVGQSLP
PSAITEIKLYSNMFASRLDLKLFIDRSRVTEVIGTEPQDLIEKTLKLVHVGQSLP
LRYAHLHLLVQVTTYRLLSRKGGWVQVSYATVVHNSRSRPHICISVNVYLIE
IYKELQISLEQVSTAKSQDSWRTALSTQETRLKVPKNTMKTKLRTNPPYPPQFIS
SPQDKLECGGLGNWRASPPASAAAPPELOPHSBSDLITYPSLSLFPYHGFPLD
SHFSSKPKMLPAKPGQPGQPEVAREFLFTMPASGBCOWHYANPLVPLVSSSPAKNP
PEPPANTAHSLVPSYEGSGLLYGVKVLRTAGSRSSHGGGWQMETEPSPFGQTCPL
SASK"
BASE COUNT 701 a 803 c 778 g 577 t
ORIGIN
Query Match 100.0%; Score 2859; DB 9; Length 2859;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACTCACTATAGGCTCGAGCGCGCGCGGCGAGGTGGGGCTCCGGGGCTCGAGCAGC 60
DB 1 ACTCACTATAGGCTCGAGCGCGCGCGGCGAGGTGGGGCTCCGGGGCTCGAGCAGC 60
QY 61 GCGGGTCTAATATGCGCGGCGCGGCGGCGGCGATGAAGAGAAGTCCAAAGATGGCGCA 120
DB 61 GCGGGTCTAATATGCGCGGCGCGGCGGCGGCGATGAAGAGAAGTCCAAAGATGGCGCA 120
QY 121 AGACGAGGAGGAGAAATGGCGAGTTTACGAGCTTGCCAGCTGCTCCCGCTGC 180
DB 121 AGACGAGGAGGAGAAATGGCGAGTTTACGAGCTTGCCAGCTGCTCCCGCTGC 180
QY 181 CGTGGCCATCACTTCCAGCTGACAAAGCGTCCATCATCCGCCTCACACAGCTACC 240
DB 181 CGTGGCCATCACTTCCAGCTGACAAAGCGTCCATCATCCGCCTCACACAGCTACC 240
QY 241 TGAAGATGCGCGCGCTCTCCCGGAGGTTTAGGAGACGCGTGGGACAGCCGAGCGCG 300
DB 241 TGAAGATGCGCGCGCTCTCCCGGAGGTTTAGGAGACGCGTGGGACAGCCGAGCGCG 300
QY 301 CCGGGCCCTGGAGCGGCTGCCAAGGAGCTGGGATCGCACTTGCTGCGACTTTGGATG 360
DB 301 CCGGGCCCTGGAGCGGCTGCCAAGGAGCTGGGATCGCACTTGCTGCGACTTTGGATG 360
QY 361 GATTGTTTGTGTAGCATCTGATGCAAAATCATGTATATATCCGAGACCGCTTCTG 420
DB 361 GATTGTTTGTGTAGCATCTGATGCAAAATCATGTATATATCCGAGACCGCTTCTG 420
QY 421 TCCATTAGGCTATCCAGCTGAGCTCAGCGGCAACAGCTATTATTATATCATCATC 480
DB 421 TCCATTAGGCTATCCAGCTGAGCTCAGCGGCAACAGCTATTATTATATCATCATC 480
QY 481 CTTCCTGACACGATGAGATGACCGCTGCTCTCAGCGGCCACCGCCGCTGCACCCACC 540

481 CTTCCTGACACGATGAGATGACCGCTGCTCTCAGCGGCCACCGCCGCTGCACCCACC 540
QY 541 TGCTCCAAAGAGTATGAGATGAGAGAGTGTCTTTCTTCGAATGAATGTCTCTTGGCGA 600
DB 541 TGCTCCAAAGAGTATGAGATGAGAGAGTGTCTTTCTTCGAATGAATGTCTCTTGGCGA 600
QY 601 AAAGAAACGCGGCTGACCTGCAGCGGATCAAAAGGTCAATCACTGAGTGGCTACTTGA 660
DB 601 AAAGAAACGCGGCTGACCTGCAGCGGATCAAAAGGTCAATCACTGAGTGGCTACTTGA 660
QY 661 AGATCAGGAGTATATGCTGGACATGTCCTGTAGACTCTCTGCTTACAGATTGTGGGCG 720
DB 661 AGATCAGGAGTATATGCTGGACATGTCCTGTAGACTCTCTGCTTACAGATTGTGGGCG 720
QY 721 TGGTGGCGCTGGGCGAGTGGCTGCCACCCAGTGCCTATCAGGAGATCAAGCTGTACAGTA 780
DB 721 TGGTGGCGCTGGGCGAGTGGCTGCCACCCAGTGCCTATCAGGAGATCAAGCTGTACAGTA 780
QY 781 ACATGTTTCATGTTTCAGGGCCAGCTTGACCTGAAGCTGATATCTCGATTCCAGGTTGA 840
DB 781 ACATGTTTCATGTTTCAGGGCCAGCTTGACCTGAAGCTGATATCTCGATTCCAGGTTGA 840
QY 841 CCAGGTGACGGGTTACGAGCCGAGACTGATCGAGAAGACCTTATACCATCATCGTGC 900
DB 841 CCAGGTGACGGGTTACGAGCCGAGACTGATCGAGAAGACCTTATACCATCATCGTGC 900
QY 901 ACGGCTGCGAGCTGTTTCCACCTCCGCTAGCAGCACACACCTCTCTGTTGGTGAAGGCCAG 960
DB 901 ACGGCTGCGAGCTGTTTCCACCTCCGCTAGCAGCACACACCTCTCTGTTGGTGAAGGCCAG 960
QY 961 TCACCAACCAAGTACTACCGGCTGCTGCTCAAGCGGGCGGCTGGGTGTGGGTGCAGAGCT 1020
DB 961 TCACCAACCAAGTACTACCGGCTGCTGCTCAAGCGGGCGGCTGGGTGTGGGTGCAGAGCT 1020
QY 1021 ACGCCACCGTGGTGCAACACAGCGCTGCTGCCGGGCCCACTGCATCGTGTGAGTGTCAATT 1080
DB 1021 ACGCCACCGTGGTGCAACACAGCGCTGCTGCCGGGCCCACTGCATCGTGTGAGTGTCAATT 1080
QY 1081 ATGTAATCAGGAGATTGAATACAGAACTTCAGCTGCTCCCTGGAGCAGCTGCTCCACTG 1140
DB 1081 ATGTAATCAGGAGATTGAATACAGAACTTCAGCTGCTCCCTGGAGCAGCTGCTCCACTG 1140
QY 1141 CCAAGTCCAGGACTCTCGGAGGACCGCTTGTCTACCTCACAAGAACTAGGAATTTAG 1200
DB 1141 CCAAGTCCAGGACTCTCGGAGGACCGCTTGTCTACCTCACAAGAACTAGGAATTTAG 1200
QY 1201 TGAACCCCAAAATACCAAGATGAAGACAAAGCTGAGAACAAACCTTACCCCCACAGC 1260
DB 1201 TGAACCCCAAAATACCAAGATGAAGACAAAGCTGAGAACAAACCTTACCCCCACAGC 1260
QY 1261 AATACAGCTCATTCCTCAATGGACAACTTGAATGGGCGGCTCGGAACTGGAGAGCCA 1320
DB 1261 AATACAGCTCATTCCTCAATGGACAACTTGAATGGGCGGCTCGGAACTGGAGAGCCA 1320
QY 1321 GTCCCGCTTCAAGCGCTGCTCTCTCCAGAACTGCAGCCCGCTCAGAAAGCAGTGAAC 1380
DB 1321 GTCCCGCTTCAAGCGCTGCTCTCTCCAGAACTGCAGCCCGCTCAGAAAGCAGTGAAC 1380
QY 1381 TTCTGTACACGCGCTTACAGCTGCTCTCTCTACCATTTGAGACACTTCCCTCTCTG 1440
DB 1381 TTCTGTGTACACGCGCTTACAGCTGCTCTCTCTCTACCATTTGAGACACTTCCCTCTCTG 1440
QY 1441 ACTCTCACTTCTTACAGCAGCAAAAGCAATGTTGCCGGCCCAAGTTCGGGCGAGCCCAAG 1500
DB 1441 ACTCTCACTTCTTACAGCAGCAAAAGCAATGTTGCCGGCCCAAGTTCGGGCGAGCCCAAG 1500
QY 1501 GATCCCGCTTGTGAGGTGGCAGCTTTTCTCTCAGCACAATGCCAGCCAGCGGTGAATGCC 1560
DB 1501 GATCCCGCTTGTGAGGTGGCAGCTTTTCTCTCAGCACAATGCCAGCCAGCGGTGAATGCC 1560
QY 1561 AGTGGCATTTATGCCAAACCCCTTAGTGCCTTAGCAGCTGCTCTCCAGCTTAAATCTCCAG 1620
DB 1561 AGTGGCATTTATGCCAAACCCCTTAGTGCCTTAGCAGCTGCTCTCCAGCTTAAATCTCCAG 1620

Db	1561	AGTGGCATTATGCCAACCCCTAGTGCCTAGCAGCTCCTCTCCAGCTAAAAATCCTCCAG	1620
QY	1621	AGCCACCGGCAACACTGTAGCCACAGCCTGGTGCCAAAGCTACGAAGTGGGTGAGGATC	1680
Db	1621	AGCCACCGGCAACACTGTAGCCACAGCCTGGTGCCAAAGCTACGAAGTGGGTGAGGATC	1680
QY	1681	TGCTCGTGGGAAAGTGGGAGGACTCGCACCGCGCGGAGCGAAGCAGCCATGCGGGTG	1740
Db	1681	TGCTCGTGGGAAAGTGGGAGGACTCGCACCGCGCGGAGCGAAGCAGCCATGCGGGTG	1740
QY	1741	GGTGGCAGATGGAGACAGAACCCCTCACGCTTTGGGCAAACTTGCCTCTTTCTGCTTCTA	1800
Db	1741	GGTGGCAGATGGAGACAGAACCCCTCACGCTTTGGGCAAACTTGCCTCTTTCTGCTTCTA	1800
QY	1801	AGTAGGCTTGTGCTTCTTGTCTCAATGCAGGTGCTCCTCGAGAGTGAGAAATGG	1860
Db	1801	AGTAGGCTTGTGCTTCTTGTCTCAATGCAGGTGCTCCTCGAGAGTGAGAAATGG	1860
QY	1861	CAGTCGTGCTGCTCGGGGACACTAGTGACAGTATAAAGGGCAAAAGGAGGATC	1920
Db	1861	CAGTCGTGCTGCTCGGGGACACTAGTGACAGTATAAAGGGCAAAAGGAGGATC	1920
QY	1921	TGGCTTTCACGTAAATCTCGGCCACATTCACCAACCAAGGGGACAGTATTTCAAAA	1980
Db	1921	TGGCTTTCACGTAAATCTCGGCCACATTCACCAACCAAGGGGACAGTATTTCAAAA	1980
QY	1981	CCAGCTCCCATGTGCTGAGAACACCCAGCTGCTTCTTTGCAAGATTCCTTTCCACT	2040
Db	1981	CCAGCTCCCATGTGCTGAGAACACCCAGCTGCTTCTTTGCAAGATTCCTTTCCACT	2040
QY	2041	CCAAACAGAGTGAATTTTGAGACAAACGGCCTATTGGCTATTTTCCCATGCCAGTTT	2100
Db	2041	CCAAACAGAGTGAATTTTGAGACAAACGGCCTATTGGCTATTTTCCCATGCCAGTTT	2100
QY	2101	GGAAGTGGGAAACATATGTTGGAAATTTTGGCTTGGGACAGAAATGCCATCACA	2160
Db	2101	GGAAGTGGGAAACATATGTTGGAAATTTTGGGCTTGGGACAGAAATGCCATCACA	2160
QY	2161	ACCCAGGCAAGAACACAAACCTCCAGGCTCAGTTTCTTCCACTGTAATATGGGGTG	2220
Db	2161	ACCCAGGCAAGAACACAAACCTCCAGGCTCAGTTTCTTCCACTGTAATATGGGGTG	2220
QY	2221	AAGCTGTATGTGCTACTCCCAAGGACACACAGTAGGAGACCTGCCTGTACATG	2280
Db	2221	AAGCTGTATGTGCTACTCCCAAGGACACACAGTAGGAGACCTGCCTGTACATG	2280
QY	2281	CTAGTTCAACAGAAAGGAATGGCCTTTACCTTCTCCTGTGGCAGGCAAGCAGATGCC	2340
Db	2281	CTAGTTCAACAGAAAGGAATGGCCTTTACCTTCTCCTGTGGCAGGCAAGCAGATGCC	2340
QY	2341	TCTGGGAGATACCGCTCAGCTCCCGAGGACGAGTACTCCTGTTTGTCTGCTGAC	2400
Db	2341	TCTGGGAGATACCGCTCAGCTCCCGAGGACGAGTACTCCTGTTTGTCTGCTGAC	2400
QY	2401	AACCCAGGCAAGAGTGGAGGTGGGAACAGAGTTTACCTGCAGGACATGATTCCCA	2460
Db	2401	AACCCAGGCAAGAGTGGAGGTGGGAACAGAGTTTACCTGCAGGACATGATTCCCA	2460
QY	2461	TTGCACGAGACCTTAATCGCGCTCAGAGTGTAGACCGAGATGGTGACAGTGCCTCAGT	2520
Db	2461	TTGCACGAGACCTTAATCGCGCTCAGAGTGTAGACCGAGATGGTGACAGTGCCTCAGT	2520
QY	2521	GCCATTAAATGTGGGTGAAGTGCATCAGGATATTGTCCCGAGGCGGGCTCAGTGG	2580
Db	2521	GCCATTAAATGTGGGTGAAGTGCATCAGGATATTGTCCCGAGGCGGGCTCAGTGG	2580
QY	2581	CTCACCTGTAATCCAGCAGCTTTGGGAGGCAAGGTGGGCGGATCAGCTAGGTCAGG	2640
Db	2581	CTCACCTGTAATCCAGCAGCTTTGGGAGGCAAGGTGGGCGGATCAGCTAGGTCAGG	2640
QY	2641	AGTTTGGGACAGCCCTGCCAAGCTGAAACCCCATCTCCACTAAAAATACAAAAATTA	2700
Db	2641	AGTTTGGGACAGCCCTGCCAAGCTGAAACCCCATCTCCACTAAAAATACAAAAATTA	2700
QY	2701	GTGGGCATGCTGGTGAGCACCTCTAATCCAGCTACTCTGGAGGCTGAGATAGGAGGAT	2760
Db	2701	GTGGGCATGCTGGTGAGCACCTCTAATCCAGCTACTCTGGAGGCTGAGATAGGAGGAT	2760
QY	2761	CACCTTGAACCCGGGAGGTGGAGGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC	2820
Db	2761	CACCTTGAACCCGGGAGGTGGAGGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC	2820
QY	2821	TGGGTAAACAGAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC	2880
Db	2821	TGGGTAAACAGAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC	2880
RESULT 2			
HSU80456			
LOCUS			
DEFINITION Human transcription factor SIM2 long form mRNA, complete cds.			
ACCESSION U80456			
VERSION U80456.1			
KEYWORDS GI:2062416			
SOURCE			
ORGANISM Homo sapiens.			
REFERENCE			
AUTHORS Chrest, R., Kudoh, J., Chen, H., Rossier, C., Chen, H., Minoshima, S., Wang, Y., Shimizu, N., and Antonarakis, S.E.			
TITLE Cloning of two human homologs of the Drosophila single-minded gene SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome chromosomal region			
JOURNAL Genome Res. 7 (6), 615-624 (1997)			
MEDLINE 97343329			
PUBMED 9199934			
REFERENCE			
AUTHORS Chrest, R., Kudoh, J., Rossier, C., Chen, H., Minoshima, S., Shimizu, N.			
TITLE Direct Submission			
JOURNAL Submitted (29-NOV-1996) Medical Genetics, University of Geneva			
FEATURES			
source			
1..3921			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="21"			
/map="21q22.2"			
93..2096			
/codon_start=1			
/product="transcription factor SIM2 long form"			
/protein_id="AAB62396.1"			
/db_xref="GI:2062417"			
/translation="MKRSKNAKTRREKNGEYELAKLLPLPSAITSQDKASIIIR LITSTYKRWAVFPEGLDAGWOPSRAGLDGVAKEGLSHLLQTLDFGVVVASDCKIM YISEFASVHLGLSQVELTNSIYIEHPSDHDEMTAVLTADHQLHLLLOEYERSE FLRMKCVLAKRNAGLTCGKYVHCQVKIKROYMLDMSLYDSCVOIVGLVAVGSQLP PSAITEIKLYSNMFMRASLDLKLIFLDSRVTEVTGEPDOLIEKLYHVHGGDVEH LRYAHLLLVKQVTKYKYLRSKRGHWVQSVATVYVHNSRSRPHCVISVNTVLTE LEYKELQSLQSVTAKSDSWRTALSTQETRKLVKPKNTKMTKURTPYPPQYS SEFQDKLSCGQLGNRASPASAAAPPELOPHSESSDLLYTPSYSLPFSYHYHGFPLD SHPSSKKKPLPAKFGQPGQCEVAREFTLTPASGECOMHWANPLVPSSSPAKNP PEPPANTARSLVPSYEAAAVRRFGEDTAPPFPSCGHYREEPALGPAAKAAAR DGAALAAAPCAPECAPTPEAPGAPQLPFVLLNLYHRLVLRPLGAAAPASGLAC APGPEAATGALRLRHPSAATSPGAPLPHYLGLASVLIITNGR"			
BASE COUNT 922 a 1137 c 998 g 864 t			
ORIGIN			
Query Match 58.1%; Score 1661.6; DB 9; Length 3921;			
Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1664; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY	1	ACTCATTATAGGCTCGAGCGCGCGCGGCGGCGGCTCGGCGGCTCGAGCAGC	60
Db	1	ACTCATTATAGGCTCGAGCGCGCGCGGCGGCGGCTCGGCGGCTCGAGCAGC	60

```
QY 61 GCCGGGCTTAATATATGCCCGGAGCGAGCGCGGATGAAGGAGAAAGTCAAGATCCGGCCA 120
Db 61 GCCGGGCTTAATATATGCCCGGAGCGAGCGCGGATGAAGGAGAAAGTCAAGATCCGGCCA 120
QY 121 AGACCAGGAGGAGAGAAATGCGAGTTTACGAGCTTGCCAAAGCTGCTCCCGCTGC 180
Db 121 AGACCAGGAGGAGAGAAATGCGAGTTTACGAGCTTGCCAAAGCTGCTCCCGCTGC 180
QY 181 CGTCGGCCATCACTTCGAGCTGGACAAAGCTCCATCATCCGCTCACACAGAGTACC 240
Db 181 CGTCGGCCATCACTTCGAGCTGGACAAAGCTCCATCATCCGCTCACACAGAGTACC 240
QY 241 TGAAGATGGCGCCGCTCTCCCGAAGTTTAGGAGCGCTGGGACAGCGGAGCGCG 300
Db 241 TGAAGATGGCGCCGCTCTCCCGAAGTTTAGGAGCGCTGGGACAGCGGAGCGCG 300
QY 301 CCGGGCCCTGACCGGCTGCGCAAGGAGCTGGGATCGCACTTGCTGCAGACTTTGGATG 360
Db 301 CCGGGCCCTGACCGGCTGCGCAAGGAGCTGGGATCGCACTTGCTGCAGACTTTGGATG 360
QY 361 GATTTGTTTGTGTAGCATCTGATGGCAAAATCATGTATATATCCGAGCGCTTCG 420
Db 361 GATTTGTTTGTGTAGCATCTGATGGCAAAATCATGTATATATCCGAGCGCTTCG 420
QY 421 TCCATTTAGGCTTATCCAGGTGGAGCTCACGGCAACAGATATTTATGAATACATCCATC 480
Db 421 TCCATTTAGGCTTATCCAGGTGGAGCTCACGGCAACAGATATTTATGAATACATCCATC 480
QY 481 CTCTGACCACGATGAGATGACCGTGTCTCACGGCCCAACAGCGCTGCACACACC 540
Db 481 CTCTGACCACGATGAGATGACCGTGTCTCACGGCCCAACAGCGCTGCACACACC 540
QY 541 TGTCTCAAGATGATGAGATGAGAGTCTCTTCTTCGATGAATGTCTTGGCGA 600
Db 541 TGTCTCAAGATGATGAGATGAGAGTCTCTTCTTCGATGAATGTCTTGGCGA 600
QY 601 AAAGGAACGGGCGCTGACCTGCAGCGGATACAAGGTATCACTGCAGTGCCTACTTGA 660
Db 601 AAAGGAACGGGCGCTGACCTGCAGCGGATACAAGGTATCACTGCAGTGCCTACTTGA 660
QY 661 AGATCAGGAGTATATGCTGGACATGCTCCTGTACGACTCCTGCTACAGATGTTGGGC 720
Db 661 AGATCAGGAGTATATGCTGGACATGCTCCTGTACGACTCCTGCTACAGATGTTGGGC 720
QY 721 TGTGGCGGCGGAGTGTCTGCCACCGAGTCCATCACCAGATCAAGCTGTACAGTA 780
Db 721 TGTGGCGGCGGAGTGTCTGCCACCGAGTCCATCACCAGATCAAGCTGTACAGTA 780
QY 781 ACATGTTTCATGTTTACGGCCAGCCTTGACCTGAAGCTGATATTCCTGGATTCCAGGTTGA 840
Db 781 ACATGTTTCATGTTTACGGCCAGCCTTGACCTGAAGCTGATATTCCTGGATTCCAGGTTGA 840
QY 841 CCGAGGTGACGGTTACGAGCCGAGGACCTGATCGAGAGACCCCTATACCATCAGTGC 900
Db 841 CCGAGGTGACGGTTACGAGCCGAGGACCTGATCGAGAGACCCCTATACCATCAGTGC 900
QY 901 ACGGCTCGAGCTGTTCCACCTCCGCTACGACACACACCTCCTGTTGGTGAAGGGCCAGG 960
Db 901 ACGGCTCGAGCTGTTCCACCTCCGCTACGACACACACCTCCTGTTGGTGAAGGGCCAGG 960
QY 961 TCACCACCAAGTACTACCGCTGCTGCCAAGGGGGCGGCTGGTGTGGGTGCAGAGCT 1020
Db 961 TCACCACCAAGTACTACCGCTGCTGCCAAGGGGGCGGCTGGTGTGGGTGCAGAGCT 1020
QY 1021 ACGCACCGTGTGCACAAAGCGCTCGTCCGCGGCCCCACTGCATGCTGAGTGTCAATT 1080
Db 1021 ACGCACCGTGTGCACAAAGCGCTCGTCCGCGGCCCCACTGCATGCTGAGTGTCAATT 1080
QY 1081 ATGTACTACGAGATTAATACAGGAATCTTACGCTGCTCCGAGCAGGTGTCCACTG 1140
Db 1081 ATGTACTACGAGATTAATACAGGAATCTTACGCTGCTCCGAGCAGGTGTCCACTG 1140
```

exon

```
QY 1141 CCAAGTCCCAGGACTCCTGGAGGACCGCCTTGTCTACCTCACAAGAACTAGGAATTAG 1200
Db 1141 CCAAGTCCCAGGACTCCTGGAGGACCGCCTTGTCTACCTCACAAGAACTAGGAATTAG 1200
QY 1201 TGAACCCCAAAATACCAAGATGAACAAAGCTGAGAAACAAACCTTACCCCCACACG 1260
Db 1201 TGAACCCCAAAATACCAAGATGAACAAAGCTGAGAAACAAACCTTACCCCCACACG 1260
QY 1261 AATACAGCTCATTTCCAAATGGACAAACTGGAATGCGGCCAGCTCGGAAACTGGAGGCCA 1320
Db 1261 AATACAGCTCATTTCCAAATGGACAAACTGGAATGCGGCCAGCTCGGAAACTGGAGGCCA 1320
QY 1321 GTCCCCCTCAAGCGTGTCTCCAGCAACTGCAGCCCCACTCAGAAAGCAGTGACC 1380
Db 1321 GTCCCCCTCAAGCGTGTCTCCAGCAACTGCAGCCCCACTCAGAAAGCAGTGACC 1380
QY 1381 TTCTGTACAGCGCATCTACAGGCTGCCCTTCTTCCATCATTTATGACACTTCCCTCTGG 1440
Db 1381 TTCTGTACAGCGCATCTACAGGCTGCCCTTCTTCCATCATTTATGACACTTCCCTCTGG 1440
QY 1441 ACTCTCACTTCTTCAGCAGCAAAAGCAATGTTGCGGCCAAGTTTCGGGACGCCCAAG 1500
Db 1441 ACTCTCACTTCTTCAGCAGCAAAAGCAATGTTGCGGCCAAGTTTCGGGACGCCCAAG 1500
QY 1501 GATCCCTTGTGAGTGGCAGCTTTTCTCTGAGCAACAATGCCAGCGCGTGAATGCC 1560
Db 1501 GATCCCTTGTGAGTGGCAGCTTTTCTCTGAGCAACAATGCCAGCGCGTGAATGCC 1560
QY 1561 AGTGGCATTTATGCCAACCCCTAGTGTCTAGCAGCTCTGTCTCCAGCTTAAATCCTCCAG 1620
Db 1561 AGTGGCATTTATGCCAACCCCTAGTGTCTAGCAGCTCTGTCTCCAGCTTAAATCCTCCAG 1620
QY 1621 AGCCACCGGCGGAACACTGCTAGGACAGCGCTGGTGCACAGTACGAAG 1668
Db 1621 AGCCACCGGCGGAACACTGCTAGGACAGCGCTGGTGCACAGTACGAAG 1668
```

RESULT 3

```
AP000697 136219 bp DNA linear PRI 14-APR-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q22.2, BAC clone:KB594G10,
DEFINITION CBRI-HLCS region.
ACCESSION AP000697
VERSION AP000697.1 GI:6712194
KEYWORDS SIM2; single-minded 2 protein; HLCS; holocarboxylase synthetase.
SOURCE Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:Keio
BAC library clone:KB594G10.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Shibuya,K., Kudoh,J., Minoshima,S., Kawasaki,K., Nakatoh,E.,
Shintani,A., Asakawa,S. and Shimizu,N.
TITLE Genomic sequencing of 1.2-Mb region on human chromosome 21q22.2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 136219)
AUTHORS Shimizu,N.
TITLE Direct Submission
JOURNAL Submitted (10-NOV-1999) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nsnimizu@med.keio.ac.jp,
Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES
Location/Qualifiers
1..136219
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.2"
/clone="KB594G10"
/cell_line="FLEB14-14"
/cell_type="pre-pro-B cell"
/clone_lib="Keio BAC library"
80..104
```


exon	/note="GENSCAN, score = 3.50, comment = Internal_exon" /evidence-not_experimental 473. .576	repeat_region	/rpt_family="MIR" complement(10097. .10832) /evidence-not_experimental /rpt_family="HAL1" complement(10251. .10334)
repeat_region	/note="GENSCAN, score = 10.54, comment = Terminal_exon" complement(990. .1314) /evidence-not_experimental /rpt_family="MER7A" 1228. .1233	exon	/note="GRAIL, score = 51.000%, comment = good" /evidence-not_experimental complement(11947. .12253) /evidence-not_experimental /rpt_family="AluX" complement(12605. .13033)
exon	/note="GENSCAN, score = -0.45, comment = Poly-A_signal" /evidence-not_experimental 1341. .1361	repeat_region	/rpt_family="MLT1C" complement(13464. .13584) /note="GENSCAN, score = 2.50, comment = Internal_exon" /evidence-not_experimental complement(14406. .14623)
repeat_region	/evidence-not_experimental /rpt_family="(CA)n" complement(1434. .1439) /note="GENSCAN, score = 1.05, comment = Poly-A_signal" 1495. .1519	exon	/rpt_family="MER20" 14721. .14816 /evidence-not_experimental /rpt_family="L2" 14874. .15181
exon	/evidence-not_experimental /rpt_family="AT_rich" 2023. .2072	repeat_region	/evidence-not_experimental /rpt_family="AluJo" complement(15887. .16217) /evidence-not_experimental /rpt_family="MLT1J2" complement(16153. .16192)
repeat_region	/note="GRAIL, score = 78.000%, comment = excellent" /evidence-not_experimental 3132. .3441	exon	/note="GENSCAN, score = -0.52, comment = Internal_exon" /evidence-not_experimental complement(16355. .16468) /note="GENSCAN, score = 3.32, comment = Internal_exon" /evidence-not_experimental complement(16676. .16777)
repeat_region	/evidence-not_experimental /rpt_family="AluY" 3527. .3824	repeat_region	/evidence-not_experimental /rpt_family="FLAM_A" complement(16814. .17127) /evidence-not_experimental /rpt_family="AluJb" 17287. .17313
exon	/evidence-not_experimental /rpt_family="AluJo" complement(4324. .4348)	exon	/evidence-not_experimental /rpt_family="AT_rich" complement(17346. .17444) /note="GRAIL, score = 51.000%, comment = good" /evidence-not_experimental complement(17734. .18099)
repeat_region	/note="GENSCAN, score = -3.5, comment = Terminal_exon" /evidence-not_experimental complement(5246. .5542) /evidence-not_experimental /rpt_family="AluJb" complement(6054. .6168)	repeat_region	/evidence-not_experimental /rpt_family="L1PB1" 18100. .18167 /evidence-not_experimental /rpt_family="(TATATG)n" complement(18170. .18443)
exon	/note="GRAIL, score = 93.000%, comment = excellent" /evidence-not_experimental complement(6437. .6585) /note="GENSCAN, score = 7.23, comment = Internal_exon" /evidence-not_experimental complement(6811. .6967)	repeat_region	/rpt_family="L1PB1" 18487. .18802 /evidence-not_experimental /rpt_family="MLT2FB" 18803. .19083
repeat_region	/note="GRAIL, score = 63.000%, comment = good" /evidence-not_experimental 6829. .6960	exon	/evidence-not_experimental /rpt_family="AluSg" 19084. .19106 /evidence-not_experimental /rpt_family="(CAAA)n" 19107. .19153
exon	/evidence-not_experimental /rpt_family="MIR" complement(7095. .7134) /note="GRAIL, score = 69.000%, comment = good" /evidence-not_experimental complement(7227. .7352)	repeat_region	/evidence-not_experimental /rpt_family="MLT1A1" complement(8545. .8853) /evidence-not_experimental /rpt_family="AluY" 8895. .9109
repeat_region	/evidence-not_experimental /rpt_family="MER90" complement(7468. .7514) /note="GRAIL, score = 83.000%, comment = excellent" /evidence-not_experimental complement(7529. .7598)	exon	/evidence-not_experimental /rpt_family="MLT1A1" 7599. .7867 /evidence-not_experimental /rpt_family="AluX" complement(7868. .8208)
repeat_region	/evidence-not_experimental /rpt_family="MER90" complement(7529. .7598)	repeat_region	/evidence-not_experimental /rpt_family="MLT2FB" 19154. .19182 /evidence-not_experimental /rpt_family="(TG)n" 19183. .19196
exon	/evidence-not_experimental /rpt_family="MLT1A1" complement(8545. .8853) /evidence-not_experimental /rpt_family="AluY" 8895. .9109	repeat_region	/evidence-not_experimental /rpt_family="MLT2FB" complement(19183. .19196) /evidence-not_experimental /rpt_family="MLT2FB"
repeat_region	/evidence-not_experimental /rpt_family="MER20" 9834. .10051 /evidence-not_experimental	repeat_region	

repeat_region	19197..19230	/evidence=not_experimental	19197	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2031
repeat_region	19230..19252	/rpt_family="(TA)n"	Db	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	118922
repeat_region	19230..19252	/evidence=not_experimental	Qy	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2091
repeat_region	19231..19233	/rpt_family="(GA)n"	Db	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	118982
exon	2006..20028	/evidence=not_experimental	Qy	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2151
repeat_region	2006..20028	/evidence=not_experimental	Db	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119042
exon	20265..20328	/note="MZF, score = 50.6%"	Qy	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2271
repeat_region	20265..20328	/evidence=not_experimental	Db	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119102
repeat_region	20329..20432	/evidence=not_experimental	Qy	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119162
repeat_region	20329..20432	/evidence=not_experimental	Db	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2331
Query Match	55.7%; Score 1593.6;	DB 9; Length 136219;	Qy	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119222
Best Local Similarity	99.4%; Pred. No. 0;		Db	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2391
Matches 1599; Conservative	0; Mismatches	9; Indels	Qy	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119282
Qy	1252	CCCCACAGATACAGCTCATCCAAATGGACAACTGGAATGGCGCCAGCTCGGAACT	1311	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2451
Db	118143	CGATCGCAATACAGCTCGTTCCAAATGGACAACTGGAATGGCGCCAGCTCGGAACT	118202	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119342
Qy	1312	GGAGGCCAGTCCCTCGAAGCGTGTCTCTCCAGAACTGCGAGCCCACTCAGAAA	1371	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2511
Db	118203	GGAGGCCAGTCCCTCGAAGCGTGTCTCTCCAGAACTGCGAGCCCACTCAGAAA	118262	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119402
Qy	1372	GCAGTGACCTTGTGTACAGCCATCTACAGCTGCCCTTCTCTACCATATGACACT	1431	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2571
Db	118263	GCAGTGACCTTGTGTACAGCCATCTACAGCTGCCCTTCTCTACCATATGACACT	118322	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119462
Qy	1432	TCCTCTGACCTCTACTTCTTTCAGCAGCAAAAGCCAAATGTTGCCGGCCAAAGTTCGGGC	1491	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2631
Db	118323	TCCTCTGACCTCTACTTCTTTCAGCAGCAAAAGCCAAATGTTGCCGGCCAAAGTTCGGGC	118382	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119522
Qy	1492	AGCCCAAGGATCCCTTGTGAGTGGCGACGGTCTTCTGACCAATGCCAGCCAGCG	1551	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2691
Db	118383	AGCCCAAGGATCCCTTGTGAGTGGCGACGGTCTTCTGACCAATGCCAGCCAGCG	118442	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119582
Qy	1552	GTGATGCCAGTGGCATATGCGAACCCCTAGTGCTAGCAGCTGCTCTCCAGCTAAAA	1611	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2751
Db	118443	GTGATGCCAGTGGCATATGCGAACCCCTAGTGCTAGCAGCTGCTCTCCAGCTAAAA	118502	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119642
Qy	1612	ATCCTCCAGAGCCACCGCGGACACCTGCTAGGCACAGCCCTGCTGCAAGCTACGAGGTG	1671	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2811
Db	118503	ATCCTCCAGAGCCACCGCGGACACCTGCTAGGCACAGCCCTGCTGCAAGCTACGAGGTG	118562	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	119702
Qy	1672	GGTCAGTCTGCTGTTGGGGAAGGTGGGAGGACTGCGAGCGCGGAGCGCGGAGCGCC	1731	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2859
Db	118563	GGTCAGTCTGCTGTTGGGGAAGGTGGGAGGACTGCGAGCGCGGAGCGCGGAGCGCC	118622	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2911
Qy	1732	ATGCGGTGGGTGGCAGATGGAGACAGAACCCCTCAGCTTTGGGCAAACTTGCCCTCTTT	1791	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	2951
Db	118623	ATGCGGTGGGTGGCAGATGGAGACAGAACCCCTCAGCTTTGGGCAAACTTGCCCTCTTT	118682	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	3011
Qy	1792	CTGCTTCTAAGTAGGGCTTGTGCTTCTTCTCAATGACAGTGTCTCTCAGAGT	1851	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	3051
Db	118683	CTGCTTCTAAGTAGGGCTTGTGCTTCTTCTCAATGACAGTGTCTCTCAGAGT	118742	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	3111
Qy	1852	GAGAAATGGCAGTCTGCTGCTCGGGGACACTAGTGACAGTATATAAGGGCAAGGAAA	1911	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	3151
Db	118743	GAGAAATGGCAGTCTGCTGCTCGGGGACACTAGTGACAGTATATAAGGGCAAGGAAA	118802	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	3211
Qy	1912	CCGAGTATCTGGCCTTCCAGTAAATCTGGCCCACTTACCAACCAAGGGGACAGTGA	1971	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	3251
Db	118803	CCGAGTATCTGGCCTTCCAGTAAATCTGGCCCACTTACCAACCAAGGGGACAGTGA	118862	TTTTCAAAACCAAGCTCCCATGTGCTGAGAACACCCCAAGCTGCATTTCTTTTGCAGATTC	3311

RESULT 4	AP001726	340000 bp	DNA	linear	PRI 10-MAY-2000
LOCUS	AP001726	340000 bp	DNA	linear	PRI 10-MAY-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 21g, section 70/105.				
ACCESSION	AP001726	AL163271	BA000005		
VERSION	AP001726.1	GI:7768723			
KEYWORDS	Homo sapiens DNA.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	1 (sites)				

AUTHORS

Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, K., Nagamine, K., Mitsuyama, S., Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Rieselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Leirach, H., Reinhardt, R. and Yaspo, M. L.

The DNA sequence of human chromosome 21

Nature 405 (6784), 311-319 (2000)

20289799

REFERENCE

AUTHORS

2 (bases 1 to 340000)

Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Rieselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Leirach, H., Reinhardt, R. and Yaspo, M. L.

Direct Submission

Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Human Genome Research Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)

On May 30, 2000 this sequence version replaced gi:7717344.

The chromosome 21 mapping and sequencing consortium consisting of

* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan,

* e-mail: hattori@gsr.riken.go.jp

* URL: http://hgp.gsc.riken.go.jp/

and

* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,

* e-mail: gscj-submit@genome.imb-jena.de

* URL: http://genome.imb-jena.de/

and

* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,

* e-mail: ashimizu@db-med.keio.ac.jp

* URL: http://www.db.med.keio.ac.jp/

and

* GBF, Dept. of Genome Analysis,

* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail: info.genome@gbf.de

* URL: http://genome.gbf.de/

and

* Max-Planck Institute for Molecular Genetics,

* Innesstrasse 73, D-14195 Berlin, Germany,

* e-mail: info-chr21@molgen.mpg.de

* URL: http://chr21.rz-berlin.mpg.de/

AL163271: Submitted (10-APR-2000).

Location/Qualifiers

1..340000

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

cl..42441

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

FEATURES

source

/map="21q22.1"

/clone="KB5G11, 5' partial"

/clone_lib="Keio BAC library"

/note="Accession No. AP000694"

12310..112344

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone="KB176G8"

/clone_lib="Keio BAC library"

/note="Accession No. AP000695"

62772..226523

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone="KB1572B10"

/clone_lib="Keio BAC library"

/note="Accession No. AP000696"

184265..320526

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone="KB594G10"

/clone_lib="Keio BAC library"

/note="Accession No. AP000697"

314337..>340000

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone="KUD8F6, 3' partial"

/clone_lib="KU21-D Cosmid library"

/note="Accession No. AP000698"

complement(748..1035)

/note="Alu"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

1420..1699

/note="AluSg"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

complement(1931..2227)

/note="L1PA7"

/rpt_family="LINE/L1"

/rpt_type=DISPERSED

2313..2509

/note="AluJo"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

2737..2756

/note="(TTTA)n"

/rpt_family="Simple_repeat"

/rpt_type=TANDEM

complement(2764..2880)

/note="FLAM_C"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

complement(2944..3064)

/note="L1MA10"

/rpt_family="LINE/L1"

/rpt_type=DISPERSED

complement(3350..3512)

/note="AluB"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

3699..3840

/note="L1M4"

/rpt_family="LINE/L1"

/rpt_type=DISPERSED

complement(4408..4556)

source

source

source

source

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region


```
|||||
Db 303890 CAAAAATTAGTTGGGCATGGTGGTGAGCACCTGTATCCAGCTACTCTGGAGGCTGAGA 303949
QY 2752 TAGGAGGATCACTTGAACCCGGGAGGTGGAGTTCAGTGTAGAGTCAATCATCACTGC 2811
Db 303950 TAGGAGGATCACTTGAACCCGGGAGGTGGAGTTCAGTGTAGAGTCAATCATCACTGC 304009
QY 2812 ACTCCAGCCTGGGTAAACAGAGTGAAGTGTCTCAAAAAAAGAAAAA 2859
Db 304010 ACTCCAGCCTGGGTAAACAGAGTGAAGTGTCTCAAAAAAAGAAAAA 304057

RESULT 5
AB003185
LOCUS Homo sapiens DNA for SIM2, SIM2s, partial cds.
ACCESSION AB003185
VERSION AB003185.1 GI:2257468
KEYWORDS alternative splicing; SIM2; SIM2s; SIM2 protein.
SOURCE Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:BAC
library clone:KB594610.
ORGANISM Homo sapiens
Eukaryota; Chordata; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
REFERENCE Chrast,R., Scott,H.S., Chen,H., Kudoh,J., Rossier,C., Minoshima,S.,
AUTHORS Wang,Y., Shimizu,N. and Antonarakis,S.E.
TITLE Cloning of two human homologs of the Drosophila single-minded' gene
SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome
chromosomal region
JOURNAL Genome Res. 7 (6), 615-624 (1997)
MEDLINE 97343329
REFERENCE 2 (bases 1 to 3365)
AUTHORS Chrast,R., Scott,H.S., Chen,H., Kudoh,J., Rossier,C., Minoshima,S.,
Wang,Y., Shimizu,N. and Antonarakis,S.E.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1997) Nobuyoshi Shimizu, Keio University School
of Medicine, Department of Molecular Biology; 35 Shinanomachi,
Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@db.med.keio.ac.jp,
Tel:03-3351-2370, Fax:03-3351-2370)
FEATURES
source
Location/Qualifiers
1..3365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-22.2"
/clone="KB594610"
/cell_line="FLEB14-14"
/cell_type="pre-pro-B cell"
/clone_lib="BAC library"
join(<1..409,2938..3365)
/note="alternative splicing"
/codon_start=1
/product="SIM2"
/protein_id="BAA21489.1"
/db_xref="GI:2257469"
/translation="OYSSFQMDKLECGOLGNWRASPPASAAAPPELOPHSESSDLLYT
PSYSLPFSYHYGHFPLDHSVFSKPKMLPAKPGQSPCEVARFFLSTLPASGCGW
HYANPLVPSSSSPAKNPPEPPANTARHSLVPSYEAAPAAVRRFGEDTAPFPSCGHY
REPALGPAKAAQARDGARLARAPECCAPPTPEAPGAPQALPFVLLNHRVLA
RGPLGGAPAAASGLACAPGGPEATGALRLRHPSPAATSPPGAPLPHYLGASVIIN
Gr"
<1..546
/note="alternative splicing"
/codon_start=1
/product="SIM2s"
/protein_id="BAA21490.1"
/db_xref="GI:2257470"
/translation="OYSSFQMDKLECGOLGNWRASPPASAAAPPELOPHSESSDLLYT
PSYSLPFSYHYGHFPLDHSVFSKPKMLPAKPGQSPCEVARFFLSTLPASGCGW
HYANPLVPSSSSPAKNPPEPPANTARHSLVPSYEGSGLLLVGKVGGLRTAGSRSHGG
GWQWETPSRFQTCPLSASK"
1..409
exon

Query Match 55.6%; Score 1590.4; DB 9; Length 3365;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1594; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1260 CAATACAGCTCATTTCCAAATGGACAACCTGGAAATGCGGCAGCTCGGAACCTGGAGAGCC 1319
Db 1 CAATACAGCTCGTTCCAAATGGACAACCTGGAAATGCGGCAGCTCGGAACCTGGAGAGCC 60
QY 1320 AGTCCCCCTGCAAGCGCTGCTGCTCTCCAGAACTGCAGCCCTCACTCAAGAAAGCAGTGAC 1379
Db 61 AGTCCCCCTGCAAGCGCTGCTGCTCTCCAGAACTGCAGCCCTCACTCAAGAAAGCAGTGAC 120
QY 1380 CTTCTGTACACGCCATCCCTACAGCCTGCGCTTCTCTACCATTTATGGACACTTCCTCTG 1439
Db 121 CTTCTGTACACGCCATCCCTACAGCCTGCGCTTCTCTACCATTTATGGACACTTCCTCTG 180
QY 1440 GACTCTCACTTCTTCCAGCAGCAAAAGCAATGTTGCCGGCGCAAGTTCGGGCAGCCCA 1499
Db 181 GACTCTCACTTCTTCCAGCAGCAAAAGCAATGTTGCCGGCGCAAGTTCGGGCAGCCCA 240
QY 1500 GGATCCCTTGTGAGGTGGCAGCGCTTTTCTTGAGCACTAGCCAGCAGCGGTGAATGC 1559
Db 241 GGATCCCTTGTGAGGTGGCAGCGCTTTTCTTGAGCACTAGCCAGCAGCGGTGAATGC 300
QY 1560 CAGTGGCATTTATGCCAACCCCTAGTGCCTAGCAGCTCGTCTCAGCTTAAATTCCTCCA 1619
Db 301 CAGTGGCATTTATGCCAACCCCTAGTGCCTAGCAGCTCGTCTCAGCTTAAATTCCTCCA 360
QY 1620 GAGCCACCGGGAACACTGCTAGGCACAGCCTGTGTCAGCAAGTACGAAGTGGGTCAAGT 1679
Db 361 GAGCCACCGGGAACACTGCTAGGCACAGCCTGTGTCAGCAAGTACGAAGTGGGTCAAGT 420
QY 1680 CTGCTCTGTGGGAAAGTGGGAGGACTGCGACAGCGCGGAGCGGCAAGCAGCCATGGCGT 1739
Db 421 CTGCTCTGTGGGAAAGTGGGAGGACTGCGACAGCGCGGAGCGGCAAGCAGCCATGGCGT 480
QY 1740 GGTGGCAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTGCCTCTTTCTGCTTCT 1799
Db 481 GGTGGCAGATGGAGACAGAACCTCAGCTTTGGGCAAACTTGCCTCTTTCTGCTTCT 540
QY 1800 AAGTAGGGCTTGTGCTGCTTTCTTGTCTCAATGCAGGTGCTCTCGAGAGTGAAGATG 1859
Db 541 AAGTAGGGCTTGTGCTGCTTTCTTGTCTCAATGCAGGTGCTCTCGAGAGTGAAGATG 600
QY 1860 GCAGTCTGCTGCTCGGGGACACTAGTGACAGTATATAAGGGCAAAAGGAAACCGAGTAT 1919
Db 601 GCAGTCTGCTGCTCGGGGACACTAGTGACAGTATATAAGGGCAAAAGGAAACCGAGTAT 560
QY 1920 CTGSCCTTCACTTAATCTTGGCCACATTCACCAACCAAGGGGGAGTGTATTTTCAA 1979
Db 661 CTGSCCTTCACTTAATCTTGGCCACATTCACCAACCAAGGGGGAGTGTATTTTCAA 720
QY 1980 ACCAGCTCCCATGTGCTGAGAACACCCAGCTCATTTCTTTTGAAGATTCCTTTCCAC 2039
Db 721 ACCAGCTCCCATGTGCTGAGAACACCCAGCTCATTTCTTTTGAAGATTCCTTTCCAC 780
QY 2040 TCCAACCCAGAGTGAATATTTGAGACAAAGCGCTATTGGCTATTTCCTATGCCAGTTT 2099
Db 781 TCCAACCCAGAGTGAATATTTGAGACAAAGCGCTATTGGCTATTTCCTATGCCAGTTT 840
QY 2100 TGAAGTGGGGAAAACTATGTTGGAAATTTTGGGCTTGGGACAGAAATGCCACTCACC 2159
Db 841 TGAAGTGGGGAAAACTATGTTGGAAATTTTGGGCTTGGGACAGAAATGCCACTCACC 900
QY 2160 AAGCCAGGGGAAAGAACACAAACCCCTCCAGGCTCAGTTTCTTCACTGTAAATGGGCT 2219
```

JOURNAL
COMMENT

855 California Ave.

Palo Alto, CA 94304 USA.

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* * be preserved.	1	2573:	contig	of 2573	bp in length
*	2574	3454:	gap of	unknown	length
*	3455	6817:	contig of	3363	bp in length
*	6818	7697:	gap of	unknown	length
*	7698	12025:	contig of	4328	bp in length
*	12026	12905:	gap of	unknown	length
*	12906	20457:	contig of	7552	bp in length
*	20458	21337:	gap of	unknown	length
*	21338	30062:	contig of	8725	bp in length
*	30063	30942:	gap of	unknown	length
*	30943	40460:	contig of	9518	bp in length
*	40461	41340:	gap of	unknown	length
*	41341	47214:	contig of	5874	bp in length
*	47215	48094:	gap of	unknown	length
*	48095	61927:	contig of	13833	bp in length
*	61928	62807:	gap of	unknown	length
*	62808	80757:	contig of	17950	bp in length
*	80758	81637:	gap of	unknown	length
*	81639	120000:	contig of	38363	bp in length.

FEATURES

```

L. .12000
/organism="Homo sapiens"
/db xref="taxon:9606"

```

```

clone="Q35A7"
/clone="Q35A7"
/clone="Q39D1"
/clone="Q39D1"
/clone="Q41B8"
/clone="Q41B8"
/clone="Q41C11"
/clone="Q41C11"
/clone="Q55C11"
/clone="Q55C11"
/clone_lib="LL21NC02 flow-sorted human chromosome 21
specific cosmid library (National Gene Library Project,
Lawrence Livermore National Laboratory)"
/notes="Sequence data collected from 3kb subclones
(Genome Center)"
(vector-pbf5 (J.-B. Pan, Stanford Human
Genome Center))

```

BASE COUNT	29276 a	26482 c	26504 g	29815 t	7923 others
ORIGIN					

Query Match	55.4%;	Score 1584.8;	DB 2;	Length 120000;
Best Local Similarity	99.3%;	Pred. No. 0;		
Matches 1592; Conservative	0;	Mismatches	12;	Indels 0; Gaps 0;

Qy	1252	CCCCACAGCAATACAGCTCATTCCTCAATGGACAAACTGGAATCGCGGCCAGCTCGGAACACT	1311
	117580	CCATCGACGAATACAGCTCTTCCCAATGGACAACTGGAATCGCGGCCAGCTCGGAACACT	117639
Db	1312	GGAGAGCCAGTCGCCCTGCAAGCGCTGCTGCTCTCCAGAACTGCAGCCCCACTCAGAAA	1371
Qy	117699	CGACGACGCGCGCTCCCACTGCAGGCCCTCAGAAA	117699

Accession	Gene	Accession	Gene
1372	GCAGTGA	1372	GCAGTGA
1373	GCAGTGA	1373	GCAGTGA
1374	GCAGTGA	1374	GCAGTGA
1375	GCAGTGA	1375	GCAGTGA
1376	GCAGTGA	1376	GCAGTGA
1377	GCAGTGA	1377	GCAGTGA
1378	GCAGTGA	1378	GCAGTGA
1379	GCAGTGA	1379	GCAGTGA
1380	GCAGTGA	1380	GCAGTGA
1381	GCAGTGA	1381	GCAGTGA
1382	GCAGTGA	1382	GCAGTGA
1383	GCAGTGA	1383	GCAGTGA
1384	GCAGTGA	1384	GCAGTGA
1385	GCAGTGA	1385	GCAGTGA
1386	GCAGTGA	1386	GCAGTGA
1387	GCAGTGA	1387	GCAGTGA
1388	GCAGTGA	1388	GCAGTGA
1389	GCAGTGA	1389	GCAGTGA
1390	GCAGTGA	1390	GCAGTGA
1391	GCAGTGA	1391	GCAGTGA
1392	GCAGTGA	1392	GCAGTGA
1393	GCAGTGA	1393	GCAGTGA
1394	GCAGTGA	1394	GCAGTGA
1395	GCAGTGA	1395	GCAGTGA
1396	GCAGTGA	1396	GCAGTGA
1397	GCAGTGA	1397	GCAGTGA
1398	GCAGTGA	1398	GCAGTGA
1399	GCAGTGA	1399	GCAGTGA
1400	GCAGTGA	1400	GCAGTGA
1401	GCAGTGA	1401	GCAGTGA
1402	GCAGTGA	1402	GCAGTGA
1403	GCAGTGA	1403	GCAGTGA
1404	GCAGTGA	1404	GCAGTGA
1405	GCAGTGA	1405	GCAGTGA
1406	GCAGTGA	1406	GCAGTGA
1407	GCAGTGA	1407	GCAGTGA
1408	GCAGTGA	1408	GCAGTGA
1409	GCAGTGA	1409	GCAGTGA
1410	GCAGTGA	1410	GCAGTGA
1411	GCAGTGA	1411	GCAGTGA
1412	GCAGTGA	1412	GCAGTGA
1413	GCAGTGA	1413	GCAGTGA
1414	GCAGTGA	1414	GCAGTGA
1415	GCAGTGA	1415	GCAGTGA
1416	GCAGTGA	1416	GCAGTGA
1417	GCAGTGA	1417	GCAGTGA
1418	GCAGTGA	1418	GCAGTGA
1419	GCAGTGA	1419	GCAGTGA
1420	GCAGTGA	1420	GCAGTGA
1421	GCAGTGA	1421	GCAGTGA
1422	GCAGTGA	1422	GCAGTGA
1423	GCAGTGA	1423	GCAGTGA
1424	GCAGTGA	1424	GCAGTGA
1425	GCAGTGA	1425	GCAGTGA
1426	GCAGTGA	1426	GCAGTGA
1427	GCAGTGA	1427	GCAGTGA
1428	GCAGTGA	1428	GCAGTGA
1429	GCAGTGA	1429	GCAGTGA
1430	GCAGTGA	1430	GCAGTGA
1431	GCAGTGA	1431	GCAGTGA

1432 TCCCTCGGACTCTCACTTCTTTCAGCAGCAAAAGCCAATGTTCCCGGCCAAGTTCGGGC 1491

QY 1432 TCCCTCTGGACTCTCACTTCCTTCAGCAGCAAAAAGCAATCTTGGCGGGCAAGTTCGGGC 1491

901	DB		AACCCGAGGCAAGAAACACAAACCCTCCAGGCTCAGTTTCTTCACTGTAANAATGGGT	960
2220	QY		GAAGCTGTGATGTGCCTACTCCCAAGGACACGACACAGTAGGAGCACTGCCCTGTACAT	2279
961	DB		GAAGCTGTGATGTGCCTACTCCCAAGGACACGACACAGTAGGAGCACTGCCCTGTACAT	1020
2280	QY		GCTAGTTCAACAGAAAGGAATGGCTTTCACCTTCTCCTGTGGCAGGCAAGACAGATGTC	2339
1021	DB		GCTAGTTCAACAGAAAGGAATGGCTTTCACCTTCTCCTGTGGCAGGCAAGACAGATGTC	1080
2340	QY		CTCTCGCGAGATACCGCCAGCTCCCGAGGACGACACTGACTCTCTGTTGCTCGCTGGAC	2399
1081	DB		CTCTCGCGAGATACCGCCAGCTCCCGAGGACGACACTGACTCTCTGTTGCTCGCTGGAC	1140
2400	QY		CAACCCGAGGACAGAGTGTGGAGGTGGCAACAGAGGTTTACCTCGAGGACATGTTATCCC	2459
1141	DB		CAACCCGAGGACAGAGTGTGGAGGTGGCAACAGAGGTTTACCTCGAGGACATGTTATCCC	1200
2460	QY		ATTGCACCGAGACCTTAACCTGCCGCTCAGAGTGTAGACCGAGATGGTCAGATGCCCTGCAG	2519
1201	DB		ATTGCACCGAGACCTTAACCTGCCGCTCAGAGTGTAGACCGAGATGGTCAGATGCCCTGCAG	1260
2520	QY		TGCCATTAANAATGTGGGTGAAGGTGCATCAGGATTATGTGCCACGACGCCGGCTCAGTG	2579
1261	DB		TGCCATTAANAATGTGGGTGAAGGTGCATCAGGATTATGTGCCACGACGCCGGCTCAGTG	1320
2580	QY		GCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGCGGATCACCTGAGGTGAG	2639
1321	DB		GCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGCGGATCACCTGAGGTGAG	1380
2640	QY		GAGTTTGGACAAAGCCTGCCAACAGCTGAACCCCATCTCCACTAAAAATAACAAAAATT	2699
1381	DB		GAGTTTGGACAAAGCCTGCCAACAGCTGAACCCCATCTCCACTAAAAATAACAAAAATT	1440
2700	QY		AGTTTGGGCATGTTGTTGAGCACTGTAATCCCAGCTACTCTGGAGGCTGAGATAGAGGA	2759
1441	DB		AGTTTGGGCATGTTGTTGAGCACTGTAATCCCAGCTACTCTGGAGGCTGAGATAGAGGA	1500
2760	QY		TCACTTTGAACCCGGGAGGTGGAGTTTGCAGTAGCTAAGATCACATCACTGCACTCCAGC	2819
1501	DB		TCACTTTGAACCCGGGAGGTGGAGTTTGCAGTAGCTAAGATCACATCACTGCACTCCAGC	1560
2820	QY		CTGGGTAAACAGATGTGACTGTCTCAAAAAAATAAAAAA	2859
1561	DB		CTGGGTAAACAGATGTGACTGTCTCAAAAAAATAAAAAA	1600

RESULT 6	
HSAC0000001	linear HTG 23-SEP-1996
LOCUS	120000 bp DNA
DEFINITION	Homo sapiens clone Q55C11, *** SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION	AC000001
VERSION	AC000001.1 GI:1556444
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 120000)
REFERENCE Quackenbush, J., Fan, J.-B., Cox, D.R., Myers, R.M. and Vollrath, D.
CONTIGUOUS GENOMIC SEQUENCE SPANNING A PORTION OF THE DOWN SYNDROME
REGION OF HUMAN CHROMOSOME 21

JOURNAL
REFERENCE
Unpublished
2 (bases 1
to 120000)
T Ban Y - B. Cox, D. R. Mvers, R. M. and Vollrath, D. D.

AUTHORS Quakenbush, J.; Fan, G. D.; Conrath, J. L.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1996) Stanford Human Genome Center, 855
California Ave. Palo Alto, CA 94304, USA

REFERENCE 3 (bases 1 to 120000)
AUTHORS Quackenbush, J., Fan, J.-B., Cox, D.R., Myers, R.M. and Vollrath, D.
TITLE Direct Submission

Direct Submission

[illegible]

BASE COUNT		345 a	446 c	376 g	298 t
ORIGIN					
Query Match					
Best Local Similarity 100.0%; Score 1389; DB 9; Length 1465;					
Matches 1389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	281	GTGGGGACAGCCGAGCCGCGCGCCCTGGAGCGCTGCCAAGGAGCTGGGATCGCA	340		
DB	1	GTGGGGACAGCCGAGCCGCGCGCCCTGGAGCGGTGCCAAGGAGCTGGGATCGCA	60		
QY	341	CTTGCTGCAGACTTTGGATGATTTGTTTTGTGTAGCATCTCTGATGGCAAAATCATGTA	400		
DB	61	CTTGCTGCAGACTTTGGATGATTTGTTTTGTGTAGCATCTCTGATGGCAAAATCATGTA	120		
QY	401	TATATCCGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGAGCTCAGGGCACAG	460		
DB	121	TATATCCGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGAGCTCAGGGCACAG	180		
QY	461	TATTTATGAATACATCCATCTCTGACACGATGAGATGACCGCTGTCTCAGGGCCA	520		
DB	181	TATTTATGAATACATCCATCTCTGACACGATGAGATGACCGCTGTCTCAGGGCCA	240		
QY	521	CCAGCGCTGTCCACCAACCTGCTCCAAAGATATGAGATGAGAGTGTCTTCTTCG	580		
DB	241	CCAGCGCTGTCCACCAACCTGCTCCAAAGATATGAGATGAGAGTGTCTTCTTCG	300		
QY	581	AATGAATGTGTTGGCGAAAGAACCGGGCTGACCTGCAGCGGATACAAAGTCAAT	640		
DB	301	AATGAATGTGTTGGCGAAAGAACCGGGCTGACCTGCAGCGGATACAAAGTCAAT	360		
QY	641	CCACTGCAGTGGCTACTTGAAGATCAGGAGTATATGCTGACATGTCCTGTGACGACT	700		
DB	361	CCACTGCAGTGGCTACTTGAAGATCAGGAGTATATGCTGACATGTCCTGTGACGACT	420		
QY	701	CTGCTACAGATGTGGGCTGTGGCGTGGCCAGTCTGCTGCCACCGTGCATCAG	760		
DB	421	CTGCTACAGATGTGGGCTGTGGCGTGGCCAGTCTGCTGCCACCGTGCATCAG	480		
QY	761	CGAGATCAAGCTGTACAGTAACATGTTTCAGGGCCAGCCTTGACCTGAACCTGAT	820		
DB	481	CGAGATCAAGCTGTACAGTAACATGTTTCAGGGCCAGCCTTGACCTGAACCTGAT	540		
QY	821	ATTCTGGATTCAGGGTACCGAGTACGGGTTACGAGCGCAGGACCTGATCGAGAA	880		
DB	541	ATTCTGGATTCAGGGTACCGAGTACGGGTTACGAGCGCAGGACCTGATCGAGAA	600		
QY	881	GACCTATACATCAGTGCACGGCTGCGACGCTGTTCCACCTCCGCTACGCACACCACT	940		
DB	601	GACCTATACATCAGTGCACGGCTGCGACGCTGTTCCACCTCCGCTACGCACACCACT	660		
QY	941	CTGTGTTGGTGAAGGGCAGGTACACAGCAAGTACTACCGGCTGCTCCAAAGCGGGCG	1000		
DB	661	CTGTGTTGGTGAAGGGCAGGTACACAGCAAGTACTACCGGCTGCTCCAAAGCGGGCG	720		
QY	1001	CTGGGTGTGGTGCAGAGTACGCCACCGTGTGTCACACAGCGCTGCTCCGGGCCCA	1060		
DB	721	CTGGGTGTGGTGCAGAGTACGCCACCGTGTGTCACACAGCGCTGCTCCGGGCCCA	780		
QY	1061	CTGCATCGTGAAGTCAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA	1120		
DB	781	CTGCATCGTGAAGTCAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA	840		
QY	1121	CCTGGACAGTGTCCACTGCGCAAGTCCAGGACTCCTGGAGGACCGCTTGTCTACCTC	1180		
DB	841	CCTGGACAGTGTCCACTGCGCAAGTCCAGGACTCCTGGAGGACCGCTTGTCTACCTC	900		
QY	1181	ACAAGAACTAGGAATTTAGTGAACCCCAAAATPACCAAGATGAGCAAGCTGAGAAC	1240		
DB	901	ACAAGAACTAGGAATTTAGTGAACCCCAAAATPACCAAGATGAGCAAGCTGAGAAC	960		
KLECGOLGNWRSPAPSAAPPELOPHSESSDLLTTPSYSLPFSVHYHGFPLDHSFFS					
SKPMLPAKPGQPGPCEVAREFLSTMPASGEQWYANPLVPSSSPAKNPPEPPA					
NTARHSILVPSYEG*					
1241 AAACCCCTTACCCCCACACAGCAATACAGCTCATTTCCAAATGACAAACTGGAAATCGGCCA					
1300					
DB	961	AAACCCCTTACCCCCACACAGCAATACAGCTCATTTCCAAATGACAAACTGGAAATCGGCCA	1020		
QY	1301	GCTCGGAAACTGGAGAGCCAGTCCCCCTGCAAGCGCTGCTCTCTCCAGAACTGCGAGCC	1360		
DB	1021	GCTCGGAAACTGGAGAGCCAGTCCCCCTGCAAGCGCTGCTCTCTCCAGAACTGCGAGCC	1080		
QY	1361	CCACTTCAGAAAGCACTGACCTTCTGTACACGCCATCTTACAGCCTGCGCTTCTCTTACCA	1420		
DB	1081	CCACTTCAGAAAGCACTGACCTTCTGTACACGCCATCTTACAGCCTGCGCTTCTCTTACCA	1140		
QY	1421	TTATGGACACTTCCCTCTCTGGACTCTCACTTCTTACAGCAGCAAAAGCAAAATGTTGCCGCG	1480		
DB	1141	TTATGGACACTTCCCTCTCTGGACTCTCACTTCTTACAGCAGCAAAAGCAAAATGTTGCCGCG	1200		
QY	1481	CAAGTTCCGGGAGAGCCCAAGGATCCCTTGTGAGTGGCAGCCTTTTCTCTGAGCACAAT	1540		
DB	1201	CAAGTTCCGGGAGAGCCCAAGGATCCCTTGTGAGTGGCAGCCTTTTCTCTGAGCACAAT	1260		
QY	1541	GCCAGCCAGCGGTGAATGCCAGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTCGTC	1600		
DB	1261	GCCAGCCAGCGGTGAATGCCAGTGGCATTTATGCCAACCCCTTAGTGCCTAGCAGCTCGTC	1320		
QY	1601	TCCAGCTAAATTCCTCCAGAGCCAGCCGCGGACACTGCTAGCAGCCTGCTGTCGCAAG	1660		
DB	1321	TCCAGCTAAATTCCTCCAGAGCCAGCCGCGGACACTGCTAGCAGCCTGCTGTCGCAAG	1380		
QY	1661	CTACGAGG 1669			
DB	1381	CTACGAGG 1389			
RESULT 8					
E14065					
LOCUS					
DEFINITION					
E14065					
ACCESSION					
E14065.1					
VERSION					
G1:5708748					
KEYWORDS					
JP 1997268199-A/7.					
SOURCE					
Mus sp.					
ORGANISM					
Mus sp.					
REFERENCE					
1 (bases 1 to 3614)					
AUTHORS					
SHIMIZU, N.					
TITLE					
HUMAN OR MOUSE SIM GENE					
JOURNAL					
PATENT: JP 1997268199-A 7 14-OCT-1997;					
FUJISAWA PHARMACEUT CO LTD					
COMMENT					
OS Mus sp. (mouse)					
PN JP 1997268199-A/7					
PD 14-OCT-1997					
PF 28-MAR-1996 JP 1996074261					
PR 31-MAR-1995 JP 95P 76362, 29-JAN-1996 JP 96P 12952 PI					
SHIMIZU NOBUYOSHI					
PC C07K14/47, C12N1/21, C12N15/09, C12Q1/68//C12P21/02, (C12N1/21, PC					
C12R1:19);					
CC strandedness: Double;					
CC topology: Linear;					
CC hypothetical: No;					
CC anti-sense: No;					
FH Key					
Location/Qualifiers					
FT source					
1. .3614					
FT /organism='Mus sp.'					
FT /tissue_type='embryo'					
FT 1..503					
FT 5'UTR					
FT CDS					
FT 504..2474					
FT /product='SIM protein'					
FT 2475..3614.					
FT 3'UTR					
Location/Qualifiers					
1. .3614					
FEATURES					
source					
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
DNA encoding SIM protein.					
3614 bp					
DNA					
linear					
PAT 28-JUL-1999					

/organism="Mus sp." /db_xref="taxon:10095"		833 a	1110 c	952 g	717 t	2 others
BASE COUNT	Query Match	46.3%;	Score 1323.4;	DB 6;	Length 3614;	
ORIGIN	Best Local Similarity	88.2%;	Pred. No. 7.4e-303;	Matches 1451;	Conservative	0; Mismatches 191; Indels 3; Gaps 1;
QY	27	CGGGCAGGTGGGGCTCCGGGGCTGGAGCAGCGCGGGTCTAATATGCCGGAGCGGA	86			
Db	438	CAGCCCAAGCGGAGTCCGGGGCTGGAGCGCAGCAGGCTCAATATGCCAGAGCGGA	497			
QY	87	GGCGGATGAAGGAGAAGTCCAAGAAATCGCGCCAAAGACAGGAGGAGAGAAATGGC	146			
Db	498	GGCGGATGAAGGAGAAGTCCAAGAAATCGCGCCAAAGACAGGAGGAGAGAAATGGC	557			
QY	147	GAGTTTACGAGCTTGGCAAGTGTCCCGTGGCGTGGCCATCACTTCCAGCTGGAC	206			
Db	558	GAGTTTACGAGCTTGGCAAGTGTCCCGTGGCGTGGCCATCACTTCCAGCTGGAC	617			
QY	207	AAAGCGTCCATCATCCGCTCACCACGAGCTACCTGAAGATGGCGCGCTTCCCGGAA	266			
Db	618	AAAGCGTCCATCATCCGCTCACCACGAGCTACCTGAAGATGGCGCGCTTCCCGGAA	677			
QY	267	GGTTTAGGAGAGCGGTGGGACAGCGGCGCGCGGCGCTGGAGCGGCTGCGCAAG	326			
Db	678	GGTCTAGGAGAGCGGTGGGACAGCGGACAGGAGCGGCTCTGGACAGCTGCAAAA	737			
QY	327	GAGCTGGGATCGCACTTGTGCTGACAGACTTGGATGGATTTGTTTGTGGTGGATCTGAT	386			
Db	738	GAGCTGGGATCGCACTTGTGCTGACAGCTAGATGGATTTGTTTGTGGTGGCTCTGAT	797			
QY	387	GGCAAAATCATGTATATATCCGAGACCGCTTCTGTCCATTTAGGCTTATCCAGGTTGAG	446			
Db	798	GGCAAAATCATGTATATATCTGAGACCGCTTCTGTCCATTTAGGCTTATCCAGGTTGAG	857			
QY	447	CTCAGGCGCAACAGTATTATGATATACATCCATCTCTGACCACGATGAGATGACCGCT	506			
Db	858	CTCAGGCGCAACAGTATTATGATATACATCCATCTCTGACCACGATGAGATGACCGCT	917			
QY	507	GTCTCTACGGCCACACCGGCTGTCACACACCGCTGCTCCAAGATGATGATAGAGAGG	566			
Db	918	GTACTTACAGCCACACCGCTGTCACACACCGCTGCTCCAAGATGATGATAGAGAGG	977			
QY	567	TCGTTCTTCTTGAATGAATGTCTTGGGAAAGGAAAGGAGCGGCGCTGACCTCGACG	626			
Db	978	TCGTTCTTCTTGAATGAATGTCTTGGGAAAGGAAAGGAGCGGCGCTGACCTCGACG	1037			
QY	627	GGATACAGGTGATCCAGTGCAGTGGCTACTTGAAGATCAGGAGTATATGCTGGACATG	686			
Db	1038	GGATACAGGTGATCCAGTGCAGTGGCTACTTGAAGATCAGGAGTATATGCTGGACATG	1097			
QY	687	TCCCTGTACGACTCTCTACAGATTTGTGGGCTGGTGGCGTGGCGGCTGACGCTGCA	746			
Db	1098	TCCCTGTACGACTCTCTTACAGATCTGGGGCTGGTGGCGTGGCGGCTGACGCTACCG	1157			
QY	747	CCAGTGCCATCACCAGATCAAGCTGTACAGTAAATGTTTATGTTTACGAGCGGCTT	806			
Db	1158	CCAGTGCCATCACCAGATCAAGCTGTACAGTAAATGTTTATGTTTACGAGCGGCTT	1217			
QY	807	GACCTGAAGCTGATATTCTCGATTCAGGCTGACCGAGGTGACCGGTTACGAGCGGAG	866			
Db	1218	GACCTGAAGCTGATATTCTCGATTCAGGCTGACCGAGGTGACCGGTTACGAGCGGAG	1277			
QY	867	GACCTGATCGAAGACCTTATACCATCAGCTGCGGCTGCGAGCTTCCACCTCCG	926			
Db	1278	GACCTGATCGAAGACCTTATACCATCAGCTGCGGCTGCGAGCTTCCACCTCCG	1337			
QY	927	TACGACACACCTCTCTGTGGTGAAGGCGGAGTCCACCAAGTACTACCGCTGCTG	986			
Db	1338	TATGACACACCTCTCTGTGGTGAAGGCGGAGTCCACCAAGTACTACCGCTGCTG	1397			
QY	987	TCCAAGCGGGCGGCTGGGTGCTGGGTGCAGAGCTAGCCAGCGGTGTGCACACAGCCGC	1046			
Db	1398	TCCAAGCTGGCGGCTGGGTGCTGGGTGCAGAGCTAGCCAGCGGTGTGCACACAGCCGC	1457			
QY	1047	TCGTCCGGCGCCCACTGCATCGTGAAGTCAATTTATGTACTCAGGAGATTCAATACAG	1106			
Db	1458	TCGTCCGGCGCCCACTGCATCGTGAAGTCAATTTATGTACTCAGGAGATTCAATACAG	1517			
QY	1107	GAACTTCAGCTGCTCCCTGGAGAGGTTCCACTGCCAAGTCCAGAGCTCTCTGGAGACC	1166			
Db	1518	GAACTTCAGCTGCTCTGGAGAGGTTCCACTGCCAAGTCCAGAGCTCTCTGGAGACC	1577			
QY	1167	GCCTTGTACCTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG	1226			
Db	1578	GCCTTGTACCTCACAAGAACTAGGAAATAGTGAACCCCAAAATACCAAGATGAAG	1637			
QY	1227	ACAAAGCTGAGAACCAACCCATATCCCCACAGCAATACAGCTCTTCCAAATGGACAA	1286			
Db	1638	ACAAAGCTGAGAACCAACCCATATCCCCACAGCAATACAGCTCTTCCAAATGGACAA	1697			
QY	1287	CTGGAATGGCGGCGGCTCGGAACTGGAGAGCCAGTCCCTGCAAGCGCTGCTGCTCT	1346			
Db	1698	CTGGAGTCGAGCCAGGTGGGAACTGGAGAACTAGTCCCCCAAAATGCTGTGCTCTC	1757			
QY	1347	CCAGAACTGACGCCCGCTCAGAAAGCAGTACACCTTCTGTACAGCCCATCTACAGCTG	1406			
Db	1758	CCGGAACACAGCAGCTCCATTTCAGAAAGCCAGCGAGCTTTATATGGCGCGGCTTAC	1817			
QY	1407	CCCTTCTCTACCATTTAGGACACTTCCCTCTGAGCTCTCACTTCTTCCAGCAGCAAAAG	1466			
Db	1818	CCCTTCTCTACCATTTAGGACACTTCCCTCTGAGCTCTCACTTCTTCCAGCAGCAAAAG	1877			
QY	1467	CCAATGTTGCGGCGCAAGTTCGGGCGAGCCCAAGATCCCTGTGAGGTGGACGCTTT	1526			
Db	1878	CCAGATTTGCGGCGCAAGTTCGGGCGAGCCCAAGATCCCTGTGAGGTGGACGCTTT	1937			
QY	1527	TTCCTGAGCACAATGCCAGCGGTGAATGCCAGTGGCATTTATGCCAACCCCTCTAGT	1586			
Db	1938	TTCCTGAGCACAATGCCAGCGGTGAATGCCAGTGGCATTTATGCCAACCCCTCTAGT	1997			
QY	1587	CCTAGCAGCTGCTTCCAGCTTAAATCTCTCCAGAGC---CACCGGGAACACTGTAGG	1643			
Db	1998	CCCAGCAGCTCATCGCCAGCTTAAACCTTTCTGAGCCTTCTCTCTGTGAATGCTGCCCCG	2057			
QY	1644	CACAGCGCTGTGCCAAGCTAGGAAG	1668			
Db	2058	CATGCTCTGTGCCAAGCTATGAAG	2082			
RESULT 9						
LOCUS	D64135	3614 bp mRNA linear ROD 13-FEB-1999				
DEFINITION	Mus musculus Sim2 mRNA, complete cds.					
ACCESSION	D64135					
VERSION	D64135.1	GI:1754612				
KEYWORDS	Sim2.					
SOURCE	Mus musculus (strain:ICR x Swiss Webster) whole embryo cDNA to mRNA, clone_lib=5'-stretch cDNA library (CLONTECH).					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Mus. 1 (sites)					
AUTHORS	Yamaki, A., Noda, S., Kudoh, J., Shindoh, N., Maeda, H., Minoshima, S., Kawasaki, K., Shimizu, F. and Shimizu, N.					
TITLE	The mammalian single-minded (SIM) gene: mouse cDNA structure and					
JOURNAL	encephalic expression indicate a candidate gene for Down syndrome					
MEDLINE	Genomics 35 (1), 136-143 (1996)					
REFERENCE	96299750					
AUTHORS	2 (bases 1 to 3614)					
TITLE	Shimizu, N.					
JOURNAL	Direct Submission					
	Submitted (13-SEP-1995) Nobuyoshi Shimizu, Keio University School					
	of Medicine, Department of Molecular Biology; 35 Shinanomachi,					

Shinjuku-ku, Tokyo 160-8582, Japan
(E-mail:shimizmedb.med.keio.ac.jp, Tel:81-3-3351-2370(ex.2720)),
Fax:81-3-3351-2370)

FEATURES
source
1. .3614
Location/Qualifiers
/organism="Mus musculus"
/strain="ICR x Swiss Webster"
/db_xref="taxon:10090"
/tissue_type="whole embryo"
/clone_lib="5'-stretch cDNA library (CLONTECH)"
1. .3614
/gene="Sim2"
504. .2477
/gene="Sim2"
/standard_name="single-minded gene"
/codon_start=1
/protein_id="BAA11013.1"
/db_xref="GI:1754613"
/translation="MKBSKNAATREKENGEFYEELAKLLPLPSAITSOLDKASIIIR
LTTGYLMRAVFPGLDWDGQSPRTGFLDSVAKEGLSHLLQILDGFVVFVADSGKIM
YISATSVHLGSOVELTGNISYIEYIHPSDHEMTAVLTATHPPLHLLHLOEYIEERSF
FLRMKVLAKNAGLTCGKYKVIHCQSLKIQYMLDMSLYDSQYIIVGLVAVGQSLP
PSATEIKLHNSNMEFASLIDLKIFLDSRVTELTGYEPQDLIEKTLIHHVHGCDLPH
LRYAHLHLLVGQVTKYVRLSKLGVWVQSVATVYHNSRSRPHCIIVSNVNYLTD
VEYKELQSDQVSTSKSOESWRITLSTSOETRSKAPKNTKMKTKLRTNPYPPOYS
SFQMDKLESCQVNWRTSPPTNAVAPPEQQLHSEADLLYGPVSLPFSYVYGHFPLD
SHVSSKKPGLPAKFGPQGPCEVARFELSTLPASSECQWHCANSIVPSSPAKNL
SEPSVNAARHGLVNPYEPASAAARFCEPAPPSPSCQHYREAPALGPAKAPRQNS
RDAARALARAPECCAPPAPPAQPLPFLVLLNHRVLAARRPLGSAAPGAPEAAG
SLRPRHGPVAAAPGAPRPHYLGLASVITNGR"
BASE COUNT 833 a 1110 c 952 g 717 t 2 others
ORIGIN

Query Match 46.3%; Score 1323.4; DB 10; Length 3614;
Best Local Similarity 88.2%; Pred. No. 7.4e-303;
Matches 1451; Conservative 0; Mismatches 191; Indels 3; Gaps 1;
27 CGGGGAGGTGGGCTCCGGGCTGGAGCAGCGCGGTCTAATATATGCGCGGAGCGGA 86
438 CAGCCCAAGGGAGCTCCGGGGCTTGGAGCCGAGGAGTCTAATATGCCAGAGCGGA 497
87 GCGCGATGAAGAGAGTCCAAAGATCGGCGCAACAGCAGGAGGAGAGAAATGGC 146
498 GCGCGATGAAGAGAGTCCAAAGATCGGCGCAACAGCAGGAGGAGAGAAATGGC 557
147 GAGTTTACAGCTTCCCAAGCTGCTCCCGCTGCGCTGCGCCATCACTTCGAGCTGGAC 206
558 GAGTTTACAGCTGCGCAAGCTGCTCCCGCTGCGCTGCGCCATCACTTCGAGCTGGAC 617
207 AAAGCGTCCATCATCGCGCTCACACAGCTACCTGAAGATCGGCGCGCTTTCGCCGAA 266
618 AAAGCGTCCATCATCGCGCTCACACAGCTACCTGAAGATCGGCGCGCTTTCGCCGAA 677
267 GGTTAGGAGACGGTGGGACAGCGAGCGCGCGCGCGCTGAGCGCGCTTTCGCCGAA 326
678 GGTTAGGAGACGGTGGGACAGCGAGCGCGCGCGCGCTGAGCGCGCTTTCGCCGAA 737
327 GAGCTGGGATCGCACTTTCGAGACTTTGGATGGATTTGTTTTTGGTGGATCTGAT 386
738 GAGCTGGGATCGCACTTTCGAGACTTTGGATGGATTTGTTTTTGGTGGATCTGAT 797
387 GGCAAAATCATGATATATCGGAGACGGCTTCTGTCCATTTAGCTTATCCAGGTGGAG 446
798 GGCAAAATCATGATATATCGGAGACGGCTTCTGTCCATTTAGCTTATCCAGGTGGAG 857
447 CTCAGGCGCAACAGTATTATGATATACATCATCTTCTGACCAACAGATGATGACCGCT 506
858 CTCAGGCGCAACAGTATTATGATATACATCATCTTCTGACCAACAGATGATGACCGC 917
507 GTCTCTCAGGCGCCACAGCGCTGACACACACCTGCTCCAAAGATGATGATGACGAG 566
918 GTACTTACAGCCCAACCGCTCCACCATCACCTGCTCCAAAGAGTACGATAGACGG 977

567 TCGTCTTTCTTTCGAATGAAATGTCTTGGCGAAAGAAACGCGGCGCTGACCTGAGC 626
978 TCCTTTCTTCTTTCGAATGAAATGTCTTGGCGAAAGAAACGCGGCGCTGACCTGAGC 1037
627 GGATACAAGGTCAATCCACTGCAGTGGCTACTTTGAAGATCAGCAGTATATCTGACATG 686
1038 GGATACAAGGTCAATCCACTGCAGTGGCTACTTTGAAGATCAGCAGTATATCTGACATG 1097
687 TCCCTGTAGGACTCTCTGCTACAGATTCGTTGGGCTTGGTGGCGGCTGAGTTCGCTGCA 746
1098 TCCCTGTAGGACTCTCTGCTACAGATTCGTTGGGCTTGGTGGCGGCTGAGTTCGCTGCA 1157
747 CCCAGTGCATCACGAGATCAAGCTGTACAGTAAACATGTTTCATGTTTCAGGGCCAGCCTT 806
1158 CCCAGTGCATCACGAGATCAAGCTGTACAGTAAACATGTTTCATGTTTCAGGGCCAGCCTT 1217
807 GACCTGAAGCTCATATTCCTGATTCAGGCTGACGAGTTCAGGCTTACGAGCGCAG 866
1218 GACCTGAAGCTCATATTCCTGATTCAGGCTGACGAGTTCAGGCTTACGAGCGCAG 1277
867 GACCTGATCGAAGACCTTATACCATCATGTCATGCTGTCGATGCTGTCACACCTTCCACCTCGC 926
1278 GACCTGATCGAAGACCTTATACCATCATGTCATGCTGTCGATGCTGTCACACCTTCCACCTCGC 1337
927 TAGCACACACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTTACCGGTGCTG 986
1338 TATGCAACACCTCTCTGTTGGTGAAGGCGCAGGTACACCAAGTACTTACCGGTGCTG 1397
987 TCCAAGCGGGCGGCTGGGTGTGGTGCAGAGTACGCCACCGTGGTGCACACAGCCGC 1046
1398 TCCAAGCGGGCGGCTGGGTGTGGTGCAGAGTACGCCACCGTGGTGCACACAGCCGC 1457
1047 TCGTCCCGGCGCCACCTGCTGATGTCATATGTTACTACGAGATTTGAATACAAG 1106
1458 TCCTCCCGGCGCCACCTGCTGATGTCATATGTTACTACGAGTGTGTAATACAAG 1517
1107 GAACCTCAGCTGCTCCCTGAGCAGGTGTCACCTGCGCAAGTCCAGGAGCTCTCGAGGACC 1166
1518 GAACCTCAGCTGCTCCCTGAGCAGGTGTCACCTGCGCAAGTCCAGGAGCTCTCGAGGACC 1577
1167 GCCTTGTCTACCTCACAAGAACTAGGAAATTTAGTGAACCCCAAAATACCAAGATGAAG 1226
1578 ACCTTGTCTACCTCACAAGAACTAGGAAATTTAGTGAACCCCAAAATACCAAGATGAAG 1637
1227 ACAAGCTGAGAACAAACCTTACGCCCCACAGCAATACAGCTCATTTCCAAATGGACAAA 1286
1638 ACAAGCTGAGAACAAACCTTACGCCCCACAGCAATACAGCTCATTTCCAAATGGACAAA 1697
1287 CTGGAATGCGGCCAGCTCGGAAACTGGAGAGCCAGTCCCTGCAAGCGCTGCTGCTCCT 1346
1698 CTGGAATGCGGCCAGCTCGGAAACTGGAGAGCCAGTCCCTGCAAGCGCTGCTGCTCCT 1757
1347 CCAGACTGACAGCCCACTCAGAAAGCAGTACCTTCTGTACAGCCATCCTACAGCCTG 1406
1758 CCAGACTGACAGCCCACTCAGAAAGCAGTACCTTCTGTATGAGCCGCTACAGCCTG 1817
1407 CCCTTCTCTACCTATTTAGGACACTTCCCTCTGAGCTCTCACTTCTTACAGCAGCAAAAAG 1466
1818 CCCTTCTCTACCTATTTAGGACACTTCCCTCTGAGCTCTCACTTCTTACAGCAGCAAAAAG 1877
1467 CCAATGTTGCGGGCCAAAGTTGCGGCGCCCAAGGATCCCTTGTGAGGTGGCAGCTTT 1526
1878 CCAGGATTTGCGGGCCAAAGTTGCGGCGCCCAAGGATCCCTTGTGAGGTGGCAGCTTT 1937
1527 TTCTCTGAGCACAATGCGCAGCCAGCGGTGAATGCCAGTGGCAATATGCCAACCCCTAGTG 1586
1938 TTCTCTGAGCACAATGCGCAGCCAGCGGTGAATGCCAGTGGCAATATGCCAACCCCTAGTG 1997
1587 CCTAGCAGCTCTCTCAGCTAAAATCTCTCAGAGC---CACCAGCGCAACACTCTAGG 1643
1998 CCGAGCAGCTCATCGCAGCTAAAACCTTTCTGAGCCTTCTCTGTAATGCTGCCCGG 2057
1644 CACAGCTGTTGCCAGCTACGAAG 1668

QY	1227	ACAAAGCTGAGACAAACCTTACCCCCACAGCAATACAGCTCATTCCAAATGGACAAA	1286
Db	1697	ACAAAGCTGAGAACCAACCCATATCCCCACAGCAATACAGCTCTTCCAAATGGACAAA	1756
QY	1287	CTGGAATGCGGGCAGCTCGSAAACTGGAGAGCCAGTCCGCCCTCGAAGCGGTGCTGCTCCT	1346
Db	1757	CTGGAGTGCAGCGAGGTGGGAAACTGGAGAACTAGTCCCCCACAAATGCTGTGGCTCCC	1816
QY	1347	CGAGAAGTCAGCGCCCACTCAGAAAGCAGTGCCTTCTGTACACGGCATCCTACAGCCTG	1406
Db	1817	CCGGAACAGCAGCTCCATTCCAGAAAGCAGCGACCTTTATATGGCGCGCCCTACAGCCTC	1876
QY	1407	CCCTTCTCCTTACCATTATGAGACACTTCCCTCTGGACTCTCACTTCTCAGCAGCAAAAAG	1466
Db	1877	CCCTTCTCCTTACCATTATGAGACACTTCCCATTTGGACTCTCACGCTCTTCACAGCAAGAAG	1936
QY	1467	CCAAATGTTGGGGCCCAAGTTCCGGCAGCCCCCAAGGATCCCTTGTGAGGTGGCAGCGTTT	1526
Db	1937	CCAGGATTTGCCCGCCAAGTTTCGGCAGCCCCCAAGGATCCCCGTGTGAGGTGGCAGCGTTC	1996
QY	1527	TTCCTGAGCACAAATGCCAGCCAGCGGTGAATGCCAGTGGCAATATGCCACCCCTAGTG	1586
Db	1997	TTCCTGAGCACACTGCCAGCCAGCCAGCGGAATGCCAGTGGCACTGTGCCAACTCTCTAGTG	2056
QY	1587	CCTAGCAGCTGCTCCAGCTTAAAAATCCTCCAGAGC---CACCGGGGAACCACTGCTAGG	1643
Db	2057	CCCAGCAGCTCATGCCAGCTTAAAAACCTTTCTGAGCCTTCTCCTGTGAATGCTGCCCGG	2116
QY	1644	CACAGCCTGGTGCCAAGTACGAAG	1668
Db	2117	CATGGCTGTGGCCAAACATGAAG	2141

RESULT 11
 MMU42554
 LOCUS
 DEFINITION
 ACCESSION
 U42554
 VERSION
 U42554.1
 GI:1377817
 KEYWORDS
 SOURCE
 Mus musculus.
 Mus musculus.
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 3071)
 Moffett,P., Dayo,M., Reece,M., McCormick,M.K. and Pelletier,J.
 Characterization of msim, a murine homologue of the Drosophila sim
 transcription factor
 Genomics 35 (1), 144-155 (1996)
 2 (bases 1 to 3071)
 Moffett,P.
 Direct Submission
 Submitted (07-DEC-1995) Peter Moffett, Biochemistry, McGill
 University, 3655 Drummond St., Montreal, P.Q. H3G 1F6, Canada
 Location/Qualifiers
 1..3071
 source

CDS

229..2202
/note="single-minded protein; similar to Drosophila melanogaster sim protein, Swiss-Prot Accession Number P05709"
/codon_start=1
/product="sim transcription factor"
/protein_id="AAB19098.1"
/db_xref="GI:1377818"
/translation="MKKSKNAKTRREKENGFEYELAKLLPLPSAITSQLDKASIIIR
LTTSLKMRVFPGLGDWQQPSRTGLDPSVAKEGCSHLLQTLDFVFWASDCKIM
YISSTAVHLGSLVETGNSIYVEYIHPSDHEMTAVLTAPHLRHHLLQGVETIERSF
FLRMKVCYKHNAGLTCGKYIHCSPDLYKIRQYMLDMSLYDCIYQVLAIVAGQSLP

BASE COUNT	700 a	968 c	804 g	599 t
ORIGIN				

Query Match	46.2%	Score 1320.2	DB 10	Length 3071	
Best Local Similarity	88.1%	Pred. No. 4.2e-302			
Matches 1449	Conservative	0	Mismatches 193	Indels 3	Gaps 1
QY	27	CCGGCGAGTGGGGCTCCGGGGCTCGGAGCAGCGCGGGTCTATATGTCGCCGAGCGCGA	86		
DB	163	CAGCCCAAGCGGAGCTCCGGGGCTCGGAGCGCAGGAGTCTATATGTCGCCGAGCGCGA	222		
QY	87	GGCGCGATGAAGGAGAAGTCCAAAGATCGGGCCAAAGACCAGGAGGAGAAGAAATGGC	146		
DB	223	GGCCCGATGAAGGAGAAGTCCAAATCGGGCCAAACCCAGGAGGAGAAGAAATGGC	282		
QY	147	GAGTTTTACGAGCTTGCCAAAGCTGCTCCCGCTGCGCTCGGCCATCACCTCCAGCTGGAC	206		
DB	283	GAGTTTTACGAGCTGGCCAAAGCTGCTCCCGCTGCGCTCGGCCATCACCTCCAGCTGGAC	342		
QY	207	AAAGGGTCCATCATCCGCCTCACACGAGCTACCTGAAGATCGGCGCGCTCTCCCGGAA	266		
DB	343	AAAGGGTCCATCATCCGACTCACACGACTACCTGAAGATCGGCGCGCTCTCCCGGAA	402		
QY	267	GTTTTAGGAGACGGCTGGGACAGCGAGCGCGCGCGCGCTGGACGGGTCGCCAAG	326		
DB	403	GGTCTAGGAGACGGCTGGGACAGCGAGCAGCAGGCGCTCTGGACAGCGTGCCAAA	462		
QY	327	GAGCTGGGATCGCACTTGTGTCAGACTTTCGATGGATTTGTTTGTGTAGCATCTGAT	386		
DB	463	GAGCTGGGATCCCATTTGCTGCAGACTCTAGATGGATTTGTGCTGCTGGCTCTGAT	522		
QY	387	GGCAAAATCATGTATATATCCGAGACGGCTCTGTCCATTTAGGCTTTACCAAGTGAG	446		
DB	523	GGCAAAATCATGTATATCTGAGACCGCTTCTGTCCATTTAGGCTTTGCCAGGTGGAG	582		
QY	447	CTCACGGGCAACAGTATTTATGAATACATCATCTCTGACCACCATCAGATGACCGCT	506		
DB	583	CTCACGGCAACAGTATTTATGAATACATCATCTCTGACCACCATCAGATGACCGC	642		
QY	507	GTCCTCAGGCGCCACAGCGCTGCACACCACTGCTCCAAAGATATGAGATAGAGAG	566		
DB	643	GTACTTTACAGCCCAACCCACCGCTCCACCATCCTGCTCCAAAGATACGAGATAGAGCG	702		
QY	567	TGCTTTCTTCTCGAATGAAATGTGTCTTGGCGAAAAGAACGGGGCGCTGACCTCCAG	626		
DB	703	TCTTTCTTCTCGAATGAAGTGGCTCTTGGCAAAAGAACGAGCGCTGACATCGAT	762		
QY	627	GGATAAAGGTATCATCTGAGTGGCTACTTTGAGATCAGGAGTATATGCTGGACATG	686		
DB	763	GGATAAAGGTATCTACTGCAAGCGCTTACCTAAAGATCAGACAGTACATGCTGGACATG	822		
QY	687	TCCTCTGACAGCTCTCTGCTACCAAGATCTGGGGGTGGTGGCGGTGGGCCAGTCCGCTGCCA	746		
DB	823	TCCTCTGACAGCTCTCTGTTTACCAAGATCTGGGGGTGGTGGCGGTGGGCCAGTCCGCTACCG	882		
QY	747	CCCAGTGGCCATCACCAGATCAAGCTGTACAGTACATGTTTCATGTTTCAGGGCCAGCGTT	806		
DB	883	CCCAGGCCATCACCAGATCAAGCTTCCACAGCAACATGTTTCATGTTTCAGGGCCAGCGTT	942		
QY	807	GACCTGAAGCTGATATTCCTGATTCACAGGTGACCGAGGTGACGGGTACGAGCCGACG	866		
DB	943	GACCTGAAGCTGATATTCCTGATTCACAGGTGACAGAGCTTACGGGTATGAGCCACAA	100		
QY	867	GACCTGATCAGAGAAGACCTTATACCATCAGTGCACGGCTGGAGCGTGTCCACCTCCGC	926		

```
Db 1003 GACCTGATGAAAAAGACACTGTACCATCATGTGCTGCTGTGACACCTTCCACCTCCGA 1062
Qy 927 TACCACACACCTCTCTGTTGGTGAAGGGCAGGTACACACCAAGTACTACCGCTGCTG 986
Db 1063 TATGCACACACCTCTCTGTTGGTGAAGGGCAGGTACACACCAAGTACTACCGCTGCTG 1122
Qy 987 TCCAAGCGGGCGGCTGGGTGTGGGTGACAGGTACGCGACCGTGGTGCACACACCGCG 1046
Db 1123 TCCAAGCTGGCGGCTGGGTGTGGGTGACAGGTACGCGACCGTGGTGCACACACCGCG 1182
Qy 1047 TCGTCCCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1106
Db 1183 TCGTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1242
Qy 1107 GAAGTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1166
Db 1243 GAAGTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1302
Qy 1167 GCCTTGTCTACCTCACAAGAACTAGGAAATTAGTGAAGCCCAAAATACCAAGATGAAG 1226
Db 1303 ACCTTGTCTACCTCACAAGAACTAGGAAATTAGTGAAGCCCAAAATACCAAGATGAAG 1362
Qy 1227 ACAAGCTGAGACAAACCTTACCCGCCACAGCAATACAGCTCATTTCCAAATGGACAA 1286
Db 1363 ACAAGCTGAGACAAACCTTACCCGCCACAGCAATACAGCTCATTTCCAAATGGACAA 1422
Qy 1287 CTGGAATGCGGCGGCTGCGGAACTGGAGAGCAGTCCCTTCAAGCCGCTGCTGCTGCT 1346
Db 1423 CTGAGTGCAGCCAGGTGGGAACTGGAGAACTAGTCCCTTCAAGCCGCTGCTGCTGCT 1482
Qy 1347 CCAGAACTGGAGCCGCTGAGAAAGCTGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1406
Db 1483 CCAGAACTGGAGCCGCTGAGAAAGCTGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1542
Qy 1407 CCCTTCTCTACCTATGAGCACTTCCCTTGGACTCTGCTGCTGCTGCTGCTGCTGCTGCT 1466
Db 1543 CCCTTCTCTACCTATGAGCACTTCCCTTGGACTCTGCTGCTGCTGCTGCTGCTGCTGCT 1602
Qy 1467 CCAATGTTGCGGCGGCTGCGGAACTGGAGAGCAGTCCCTTCAAGCCGCTGCTGCTGCT 1526
Db 1603 CCAGGATTGCGGCGGCTGCGGAACTGGAGAACTAGTCCCTTCAAGCCGCTGCTGCTGCT 1662
Qy 1527 TTCTGAGCACTATGCGGCGGCTGAGTGAATGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1586
Db 1663 TTCTGAGCACTATGCGGCGGCTGAGTGAATGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1722
Qy 1587 CTTAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1643
Db 1723 CCAGCAGCTATGCGGCGGCTGAGTGAATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1782
Qy 1644 CACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
Db 1783 CATGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1807
```

RESULT 12

```
MMU040576
LOCUS
DEFINITION Mus musculus single-minded 2 (Sim2) mRNA, linear ROD 02-MAR-1996
ACCESSION U40576
VERSION U40576.1 GI:1213282
KEYWORDS house mouse strain=Swiss Webster.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3963)
AUTHORS Fan, C.-M. and Tessier-Lavigne, M.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1995) Chen-Ming Fan, Anatomy, UC, San Francisco,
513 Parnassus Avenue, Rm S1479, San Francisco, CA 94143-0452, USA
FEATURES
Location/Qualifiers
1..3963
source
```

gene

CDS

```
/organism="Mus musculus"
/strain="Swiss Webster"
/db_xref="taxon:10090"
/chromosome="16"
/dev_stage="embryonic brain (E11.5)"
/notes="equivalent to the human single-minded (slmx) gene isolated by exon trapping of the Down Syndrome critical region DNA YACs"
1..3963
/gene="Sim2"
889..3117
/gene="Sim2"
/standard_name="single-minded 2"
/notes="Presumably a DNA binding protein; a homolog of Drosophila single-minded gene; bHLH-PAS homolog"
/codon_start=1
/product="SIM2"
/protein_id="AA91202.1"
/db_xref="GI:1213283"
/translation="MKESKNAATKREKGEFELAKLPLSAITSQLDKASIR
LTSYLKMRVPEGLDQWQSRTPGLDSVAKELGSHLQTLQDFVVFVADSGKIM
YISETASVHLGSLQVELTNSIYKIHPSDHDHMTAVLTAHPHLHLLLOEYERSE
FURMKCVLAKRNLGTCGYKIHGSGYKIQYMLDMSLYDSCYQVGLVAGSGLP
PSAITEIKLHNSNEMERASLDLKLIFLDSRVTELTGYEPQDLIEKTLHHVHGCDTFH
LRYAHLHLKQCVYTKYRLSKLGGWVVSATVYVHNSRSPHCIVSNVYVLTLD
VGKELQLSDQVSTKSOESWRTTLLSTQETRSKAPKMTKMLKLNQVSPSTAPS
KWNNSAARWETGELVPPQMLPRNSSIQKATLYGPPYSLPFSYHGFPLDSHV
FSSKPGPLPAKEQPGCEVARFELSTLPASSBCOWHANSLVPSYSSPAKNSLP
SCVNAARHGLVNPYAPASAAARFEDPARPSPSCGHVREPAKLVPAKLVARRPRA
RWMSAPPEPHGPAOLPVLLNHRVLAARRPLGSAAPGAPAGSLRPRHPGVAAS
APCPALPLGLRHGHQVTRQSRAGRPGRSHQAFPSGSSFTPKVQMGRTVYA
RYKLYITETIILIELRLDFYRCWCRLGRLSWEMGRGTAHSCPRAGFHGAHPGV
NL"
```

BASE COUNT	902 a	1237 c	1064 g	760 t
ORIGIN				
Query Match	44.0%	Score 1257.2	DB 10	Length 3963
Best Local Similarity	87.4%	Pred. No. 3.6e-287		
Matches 1438	Conservative 0	Mismatches 193	Indels 15	Gaps 5
QY	27	CCGGCAGGTGGGGCTCCGGCGCTGGAGCAGCGCGGTCTAATATATGCCCGGAGCGGA	86	
DB	824	CAGCCCAAGCGGACTCCGGCGCTGGAGCGCAGGAGTCTATATGCCAGAGCGGA	883	
QY	87	GCGCGATGAGGAGAAATCCAAAGATGCGGCGCAAGCAGAGGAGGAGAAATGCG	146	
DB	884	GG-GCGATGAGGAGAAATCCAAAGATGCGGCGCAAGCAGAGGAGGAGAAATGCG	942	
QY	147	GAGTTTACGAGCTTGCAGAGCTGCTCCCGCTGCGCGCTGCGCATCACCCTGCGAGTGGAC	206	
DB	943	GAGTTTACGAGCTGCGCAAGCTGCTCCCGCTGCGCGCTGCGCATCACCCTGCGAGTGGAC	1002	
QY	207	AAAGCGTCCATCATCCGCTCACCACGAGTACCTGGAAGATGCGCGCGCTCTTCCCGGAA	266	
DB	1003	AAAGCGTCCATCATCCGCTCACCACGAGTACCTGGAAGATGCGCGCGCTCTTCCCGGAA	1062	
QY	267	GTTTAGGAGAGCGCTGGGAGACAGCGCGCGCGCGCGCGCTGGAGCGGTGCGCAG	326	
DB	1063	GTTTAGGAGAGCGCTGGGAGACAGCGCGCGCGCGCGCGCTGGAGCGGTGCGCAG	1122	
QY	327	GAGCTGGGATGCGACTTGTGTCAGACTTTGGATGATTTGTTTGTGCTAGCATCTCAT	386	
DB	1123	GAGCTGGGATGCGACTTGTGTCAGACTTTGGATGATTTGTTTGTGCTAGCATCTCAT	1182	
QY	387	GGCAAAATCATGTATATATCCGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGGAG	446	
DB	1183	GGCAAAATCATGTATATATCCGAGACCGCTTCTGTCATTTAGGCTTATCCAGGTGGAG	1242	
QY	447	CTCAGGCGCACAGTATTTATGATACATCATCTCTGACACGAGTACATGACCGCT	506	
DB	1243	CTCAGGCGCACAGTATTTATGATACATCATCTCTGACACGAGTACATGACCGCT	1302	
QY	507	GTCCTACGGCGCCACCGCTGCACACCACTGCTGCCAAGAGTATGATAGATAGAGG	566	

Mon Jul 21 15:46:54 2003

us-09-923-684-2.rge

Db	983	AACCTCAGGACCTGATTGAGAGACTCTGTACACCATGTGCACGGCTCGGACACCTCC	1042
Qy	919	ACCTCCGCTACGACACACCTCCTCTGTGTGAAGGCCAGGTACACCAAGTACTACC	978
Db	1043	ACCTGGCGCTGGCGGACCATTTGCTGCTGTGAAGGACAGGTGACCAAGTACTACA	1102
Qy	979	GGCTGCTGTCCAAAGCGGGGGTGGTGTGGGTGAGAGCTACGCCACCGTGGTGACA	1038
Db	1103	GGTTCCTGGGAAACACGCGGGTGGTATGGGTGCAAGCTACGGGACATCGTGACA	1162
Qy	1039	ACAGCCCTCGTCCCGGCCCACTGCTGATGCTGATGCTCAATTATGTACTACGGGATTG	1098
Db	1163	ACAGTCGCTCTCCAGGCCACATGTATCGTCAGCGTCACTATGTCTCTCAGACACAG	1222
Qy	1099	AATACAAAGGAACTTCAGCTGTCCCTGGAGCAGGTGTCCACTGCTCAAGTCCAGGACTCT	1158
Db	1223	AATACAAAGGCTGCAGCTCTCCCTGGATCAGATCTCAGCCTCAA---ACCAGCCTTCT	1279
Qy	1159	GGAGGACCGCTTGTCTACCTACAGAAACTAGGAATTAGTGAACCCAAATACCA	1218
Db	1280	CCTATACGACGAGCTCCACCCCACTGATGCTGACAAAGAGGGGCCCAATCCCGGC	1339
Qy	1219	AGATGAAGACAAAGCTGAGAACAAACCTTACCCCCACACAGCAATACAGCTCATTCAAA	1278
Db	1340	TCTCCAGCTCAAGTCAAAATCCAGGACTTCCCATACCTCAGTATTCGGGATTCACA	1399
Qy	1279	TGGACAAACTGGAATGGCGCCAGCTCGGAAACTGGAGAGCCAGTCCCTTGCAGCGCTG	1338
Db	1400	CAGAAAGATCGGAATCTGATCATGACAGCCAGTGGGGCGAAGTCCCTTGACCGACACG	1459
Qy	1339	CTGCTCTCCAGAACTGCAGCCCACTCAGAAAGCAGTGCCTTCTGTACACGCCATCCT	1398
Db	1460	CCTCTCCGCGCTTCTGGACCCCGCGGATAGGCTGGCTCCACACAGCGCATCGTGG	1519
Qy	1399	ACAGCCTGCCCTTCTCTTACCATTTATGGACACTTCCCTCTGGACTCTCACTTTTCAGCA	1458
Db	1520	CCTACAGACAGTTTTCGGACCGCAGCTCTCTCTGTATGGCTTTGCGCTTGACCACTCGA	1579
Qy	1459	GCAAAAGCCAATGTTCCGGCCCAAGTTCCGGCAGCCCAAGGATCCCTTGTGAGGTGG	1518
Db	1580	GGCTGGTGGAGAGAGGCAATTTCCATACCCAGGCTGTGAAGGAGCGCGATGTGAGGCAG	1639
Qy	1519	CACGCTTTTCTCTGAGCACAATGCCAGCCAGCGGTGAATGCCAGTGGCAATATGCCAAC	1578
Db	1640	GCAGGTACTTCTGGAGCGCGCAGCGCGGAGGAGCCCTGTGGGGCTCTCGCGCAG	1699
Qy	1579	CCCTAGTGCCTAGCAGCTGCTCCAG	1605
Db	1700	CCTTGCCCTGACAAAGGCTCCCCAG	1726

Search completed: July 17, 2003, 19:54:13
Job time : 7282 secs